

**THESE DE DOCTORAT
DE L'ECOLE NORMALE SUPERIEURE DE CACHAN**

Présentée par

Monsieur David PHILIPONA

**pour obtenir le grade de
DOCTEUR DE L'ECOLE NORMALE SUPERIEURE DE CACHAN**

Domaine :
MATHEMATIQUES

Sujet de la thèse :

Développement d'un cadre mathématique pour une théorie sensorimotrice de l'expérience sensorielle.

Thèse présentée et soutenue à Cachan le 14 janvier 2008 devant le jury composé de :

Daniel BENNEQUIN
Sidney FRANKEL
Jean-Michel MOREL
Jean PETITOT
Jean-Pierre NADAL
Kevin O'Regan

Professeur – Paris 7
Chercheur – invité Paris 7
Professeur – ENS Cachan
Directeur d'Etudes – EHESS
Directeur de recherche – CNRS
Directeur de recherche – CNRS

Rapporteur
Rapporteur

Directeur de thèse
Co-directeur de thèse

Ce que l'on ne peut dire, il faut le taire.

Ludwig Wittgenstein, Tractatus logico-philosophicus

Résumé

De nombreuses caractéristiques du système visuel peuvent paraître surprenantes, au regard de notre expérience perceptive. Un simple exemple : la répartition des différents photorécepteurs au travers de notre rétine est très irrégulière, en type et en densité, et pourtant nous n'avons pas conscience de différences de résolution spatiale et chromatique dans la scène visuelle que nous percevons.

Les explications les plus communes, pour réconcilier de tels faits apparemment contradictoires, reposent sur l'idée que l'expérience visuelle est "produite" d'une manière ou d'une autre par l'activation de représentations corticales, divers mécanismes neuronaux correctifs opérant entre la stimulation sensorielle, en l'occurrence rétinienne, et ces représentations.

Il existe également des approches alternatives. Selon l'une d'elles, l'expérience visuelle n'est pas l'activation de représentations, mais la capacité de s'engager dans une certaine structure de fonctionnements vis-à-vis de notre environnement. De ce point de vue, nous n'avons naturellement pas conscience des caractéristiques de notre stimulation rétinienne, ni d'ailleurs d'éventuelles représentations corticales, puisque nous ne pouvons pas nous engager, vis-à-vis de ces choses, dans la structure de fonctionnements désignée par l'expression "voir ces choses", ou plus généralement "les ressentir".

Cette thèse présente une formalisation mathématique élémentaire visant à explorer une telle approche. Elle se focalise sur certaines contraintes fondamentales imposées à notre structure de fonctionnement : les contraintes imposées par les dépendances existant, dans une variété d'environnements, entre nos commandes motrices et nos stimulations sensorielles.

Une analyse des propriétés de symétrie et d'invariance de la variété de dépendances induite par notre système sensorimoteur et notre environnement montre qu'un

ensemble d'affirmations courantes sur notre expérience perceptive peuvent y trouver une explication. En particulier, il apparaît que les similarités qui peuvent être établies entre différentes dépendances, au travers d'adaptations sensorielles et motrices adéquates, sont cohérentes avec notre discours sur l'espace et les couleurs.

Le travail réalisé sur l'espace s'inspire, dans un cadre conceptuel différent, de l'argumentation développée par Poincaré sur ce qu'il nommait "l'espace sensible". Selon cette argumentation, la spatialité est un système de correspondances entre les modifications de dépendance compensables par un répertoire d'adaptations motrices. Le travail réalisé sur les couleurs, quant à lui, montre que la catégorisation des couleurs observée au travers d'études anthropologiques et psychophysiques coïncide étroitement avec les propriétés d'invariance des dépendances associées.

Organisation du manuscrit

Le manuscrit est organisé comme suit. Une première partie propose une analyse du débat scientifique autour de l'expérience perceptive, au travers de différents travaux et théories classiques en neurosciences et en psychologie de la perception. Cette analyse permet de donner une idée des difficultés soulevées par le cadre explicatif habituellement considéré pour rendre compte des données sur les sensations, et de légitimer l'intérêt d'en explorer de nouvelles.

Une seconde partie se concentre sur la notion de spatialité de l'expérience perceptive, en particulier visuelle. Elle rappelle la longue tradition d'approches sensorimotrices de l'expérience de spatialité, pour introduire les articles que nous avons rédigés sur le thème de l'espace, dont deux sont publiés et deux autres encore en préparation. Ils sont accompagnés d'une analyse critique de l'approche générale, et de propositions d'évolutions. La partie se termine en pointant ce qui paraît être des ambiguïtés dans l'argumentation informelle de Poincaré vis-à-vis de l'espace sensible.

La troisième et dernière partie du manuscrit se concentre sur la notion de structure perceptive des couleurs. Elle rappelle le modèle dichromatique des propriétés de réflexion des surfaces, et montre que ce modèle induit une distinction entre différentes propriétés chromatiques sur la base de symétries dans l'ensemble des dépendances sensorimotrices associées. Cette analyse permet d'introduire un article que nous avons publié sur la coïncidence entre les données empiriques sur la catégorisation des cou-

leurs et les propriétés de réflexion des surfaces. Elle se termine par une annexe rappelant l'histoire de l'étude des couleurs, afin de mettre cet article en perspective, et de montrer les nombreux revirements qui ont eu lieu tout au long de cette histoire vis-à-vis de la conception même de ce qu'est la couleur, et de l'interprétation des données collectées sur ce sujet.

Les éléments portant les idées les plus originales de la thèse sont : le dialogue qui sert de fil conducteur à la première partie, et les articles inclus dans le manuscrit. Pour un aperçu synthétique du coeur de la thèse, il est donc possible de se restreindre au dialogue, et aux sections 6.1 (article sur l'espace dans le cas de dépendances associatives), 6.5 (reformulation de l'approche sur l'espace), 8 (premier article, cas de dépendances différentielles), et 9.2 (article sur les couleurs).

Remerciements

Je tiens à remercier en premier lieu Kevin O'Regan et Jean-Pierre Nadal pour l'encadrement de cette thèse : Kevin pour son enthousiasme inlassable, son acharnement à poursuivre jusqu'au bout ses idées, et son soutien permanent, Jean-Pierre pour ses conseils incisifs et sa confiance réconfortante. De ces années de thèse me restera le souvenir d'une aventure intellectuelle, et j'ai été privilégié de pouvoir profiter de leur expérience au cours de cette aventure.

Je tiens également à remercier tout particulièrement les rapporteurs, Daniel Bennequin et Sidney Frankel, ainsi que les autres membres du jury, Jean-Michel Morel et Jean Petitot, qui ont très aimablement accepté de se pencher sur une thèse multidisciplinaire, exercice toujours délicat. Je souhaite d'autre part les remercier pour les discussions que nous avons pu entretenir, pendant la thèse ou à propos du manuscrit. Elles ont été pour moi une source de réflexion et d'inspiration de grande valeur.

Je remercie enfin toutes les personnes qui ont fait de ces années de thèse ce qu'elles ont été, au niveau recherche comme au niveau personnel. Il serait absurde de s'essayer à une liste exhaustive, et j'espère que ceux dont le nom n'est pas sur cette page ne prendront cela ni comme un oubli d'eux ni comme une ingratitude : devant l'injustice de l'exercice, je me suis simplement résolu à un balayage impressionniste... Je remercie tous les membres du Sony CSL, en particulier l'équipe neurosciences dont je faisais partie – Angelo Arleo, Michele Bezzi, Christian Boucheny, Olivier Coenen –, tous les membres du LPP, chercheurs et doctorants – Ed Cooke, Thérèse Collins, Hélène Gauchou, Frédéric Glanois, Victoria Medina, Paul Reeve –, les membres du réseau d'excellence Enactive et du projet européen Cosy avec qui j'ai été en relation, et bien sûr le groupe des buveurs métaphysiques – Amandine Alfonso, Malika Auvray, Cedric Coussinet, Régis Jaulin, Charles Lenay.

Pour terminer, je remercie plus généralement l'ensemble de mes amis de promos, du Jura, de rencontres diverses. Sans eux tous, la vie n'aurait pas la même saveur.

A Nicolas, dont les sourires brillent dans ma nuit.

Paris, le 14 janvier 2008

Table des matières

Résumé	i
Remerciements	v
I Science et conscience	1
1 La notion d'expérience sensorielle	3
1.1 Le retour en grâce des interrogations sur la conscience	5
1.2 Que doit expliquer une théorie de l'expérience sensorielle ?	8
2 Expérience sensorielle et système nerveux	13
2.1 Le siège de l'esprit	15
2.2 Les propriétés "paradoxaes" du système visuel	19
2.3 L'expérience sans domicile	25
3 Retour au réalisme ?	29
3.1 La phénoménologie de la conduite	31
3.2 Toile de fond de l'approche sensorimotrice	35
4 Perspectives sur les dialogues	39
II L'espace de l'expérience	47
5 Espace réel, espace sensorimoteur	49
5.1 Les fondements de la géométrie	50
5.2 La tradition sensorimotrice de l'espace	52

5.3	Différents types de dépendances sensorimotrices	56
6	Dépendances sensorimotrices associatives	61
6.1	Article : Is There Something Out There ?...	63
6.2	Article : Perception of the structure	85
6.3	Compléments aux articles	94
6.4	Critique des articles	97
6.5	Reformulation de l'approche	100
7	L'argumentation de Poincaré	113
7.1	Des fondements de la géométrie	113
7.2	Les difficultés d'interprétation de l'argumentation	118
8	Articles en préparation : dépendances différentielles	121
III	Les couleurs de l'expérience	155
9	La structure perceptive des couleurs	157
9.1	Une approche sensorimotrice	157
9.2	Article : Color naming...	162
A	L'étude des couleurs	177
A.1	Les balbutiements d'une théorie de la couleur	178
A.2	Trichromaticité et opposition	183
A.3	L'étude des capacités visuelles	187
A.4	Quelques points plus récents	189
	Conclusion	195
	Bibliographie	197

Première partie

Science et conscience

Chapitre 1

La notion d'expérience sensorielle

Un photographe, appareil numérique en bandoulière, fait un bout de chemin dans la campagne avec un chercheur.

Φ La lumière est exceptionnelle. Toutes ces couleurs, ces textures... [Il s'arrête sans cesse pour prendre des photos. Dans les champs qui les entourent, quelques vaches redressent la tête, modulant un concert de cloches, pour inspecter les deux promeneurs]. Nous ne sommes qu'en avril, et pourtant on dirait l'été ! [il tourne son visage vers le soleil, les yeux fermés, pour profiter de la chaleur] Tu travailles sur les sensations, toi, c'est ça ?

Ψ [très sérieux] Oui. Je cherche à comprendre par exemple pourquoi nous disons que voir ça [il montre le paysage], c'est très différent d'entendre ça [il siffle, en dressant l'index à côté de son oreille].

Φ [regardant l'autre, déconcerté] Mais nous le disons parce que *c'est* très différent !

Ψ Quoi, qu'est-ce qui est différent ?

Φ [il réfléchit] Et bien, ce que je vois ressemble à cette image [il attrape son appareil, prend une photo du paysage, puis met l'écran sous le nez de son compagnon] : c'est spatialement étendu, détaillé, coloré. Ce que j'entends [il émet à son tour un sifflement] est aigu ou grave, ça ressemble au son d'une cloche ou d'un moteur de voiture. Ça n'a aucun rapport !

- Ψ C'est ça, je cherche à comprendre l'origine de ce genre d'affirmations.
- Φ Mais enfin c'est très simple : voir et entendre sont des expériences différentes, j'en ai conscience, directement. C'est ça percevoir !
- Ψ [les yeux plissés] L'expérience ? De quoi est-ce que tu parles ?
- Φ [embarrassé] Et bien... [le visage soudain illuminé] par exemple, l'expérience visuelle, c'est l'image comme ça [il pointe l'écran de l'appareil photographique] qui m'apparaît quand j'ouvre les yeux. C'est incroyable qu'une image comme ça m'apparaisse quand j'ouvre les yeux !
- Ψ Quand tes paupières se soulèvent, ton organisme peut soudain s'engager dans une *structure de fonctionnements* : reconnaissance de stimuli, exploration visuelle, détection de changements, planification, discours, etc. Rien n'apparaît, tu sais.
- Φ Rien n'apparaît *vraiment*, bien sûr ! L'expérience visuelle, c'est l'image qui m'apparaît, à moi – l'image qui apparaît... dans mon esprit. D'ailleurs, peut-être qu'elle est différente de celle qui apparaît dans le tien, peut-être que leurs couleurs sont inversées ! [d'un air entendu] C'est un mystère très profond. C'est ça que tu cherches à comprendre ?
- Ψ [sceptique] Enfin, l'esprit n'est pas un lieu. Quand tu dis "X est dans mon esprit", tu exprimes simplement le fait de pouvoir t'engager dans un ensemble de fonctionnements en rapport avec X, ça ne veut pas dire que X est dans un endroit nommé "esprit". C'est cette image [il montre l'écran de l'appareil] qui est comme ça [il montre le paysage] vis-à-vis des fonctionnements dans lesquels tu peux t'engager, pas une chose dans un endroit mystérieux qui est comme ça [il montre l'écran].
- Φ [les sourcils froncés] Tu m'embrouilles, avec ton histoire. Mon problème est simple, quand je soulève les paupières, une image m'apparaît, à moi : c'est une expérience privée.
- Ψ Voyons, c'est *comme si* une image t'apparaissait soudain : la structure des fonctionnements dans laquelle tu peux t'engager évolue presque de la même manière que si tu étais dans l'obscurité et qu'une image apparaissait soudain. Et ce n'est pas *comme si* une image m'apparaissait à moi aussi parce que, quand *tes* paupières se soulèvent, cela modifie la structure des fonctionnements dans laquelle *ton* organisme peut s'engager, mais pas celle dans laquelle *mon* organisme peut

1.1. LE RETOUR EN GRÂCE DES INTERROGATIONS SUR LA CONSCIENCE⁵

s'engager.

Φ [il s'emporte, agacé] Tu ne comprends rien du tout. L'expérience visuelle, c'est *l'image-vue* que j'ai en plus de n'importe quel automate capable de s'engager dans les mêmes fonctionnements que moi, mais qui en fait n'aurait pas d'expérience. Moi, je vois vraiment, je ne fais pas juste semblant. C'est cette image-vue, ce phénomène psychophysique incompréhensible, qu'il faut expliquer !

Ψ Bien sûr que tu vois vraiment. Je prétends seulement que, selon la manière dont nous utilisons le langage ensemble tous les jours, le complément d'objet de l'expression "je vois", c'est ça [il montre le paysage]. Il ne faut pas mépriser cet usage originel du langage, c'est sur lui que tu t'appuies alors même que tu dis que ce n'est pas ce paysage que nous voyons, mais une image-vue dans notre esprit. L'expression "image-vue" n'a pas de sens sans référence à l'usage courant, contextuel, de l'expression "voir une image". [Il se tait un instant] Les images-vues sont juste une manière de parler, des "comme si", il n'y a pas de phénomène psychophysique à expliquer.

Φ Mais pas du tout, il faut expliquer les images-vues détaillées et colorées !

Ψ [las] Une image détaillée et colorée, au sens concret du discours, c'est ce qu'on pointe comme étant une image détaillée et colorée [il montre l'écran de l'appareil] ; et les images au sens figuré du discours, les images-vues, ce sont juste des équivalences de situations vis-à-vis de certains fonctionnements, *ce ne sont pas des choses*. A mon avis, ce ne sont pas les images-vues et les sons-entendus qu'il faut expliquer, mais ta capacité pratique de distinguer et comparer différentes structures de fonctionnements dans lesquels tu peux te trouver engagé, parce que c'est là que se trouve l'origine de ton discours sur l'expérience.

1.1 Le retour en grâce des interrogations sur la conscience

A la fin du XIX^e siècle, Wilhelm Wundt cherchait à établir les bases d'une "psychologie physiologique" sur l'idée que la physiologie étudie "la vie extérieure" de notre organisme, alors que "dans la psychologie, l'homme se voit du dedans". Sa définition des sensations visait à établir une science des objets de cette observation intérieure, sur le modèle atomiste de la science vedette de l'époque, la chimie (Wundt, 1886) :

[...] nous donnerons le nom de *sensations* à ces états de notre conscience qu'il est impossible de décomposer en des éléments plus simples. [...] Des sensations différentes se distinguent par leur *qualité* ; ou bien, si celle-ci est analogue, leur *intensité* peut ne pas être la même.

Un demi siècle plus tard, en 1913, John Watson dresse le constat des errances d'une science échafaudée sur ce genre d'affirmations concernant les sensations. De telles affirmations, par le concept même "d'observation intérieure", n'ont que le crédit qu'on veut bien apporter à leurs auteurs (Watson, 1913) :

There is no longer any guarantee that we all mean the same thing when we use the terms now current in psychology. Take the case of sensation. A sensation is defined in terms of its attributes. One psychologist will state with readiness that the attributes of a visual sensation are quality, extension, duration, and intensity. Another will add clearness. Still another that of order. I doubt if any one psychologist can draw up a set of statements describing what he means by sensation which will be agreed to by three other psychologists of different training.

Exit la psychologie introspectionniste : par définition, il ne peut y avoir de science, *il ne peut y avoir de rationalité partagée*, que concernant un phénomène public.

Aussi radicale qu'elle puisse paraître, la critique devait avoir une certaine pertinence, puisqu'elle a durablement jeté l'opprobre sur l'étude de la conscience, et a lancé à l'époque la plupart des scientifiques sur la voie étroite du behaviorisme.

* * *

Plusieurs décennies plus tard, avec les progrès des neurosciences, des techniques d'imagerie cérébrale et le développement de la psychologie cognitive, l'étude de la conscience est finalement revenue au premier plan de la scène scientifique. Le magazine *Science* a confirmé en 2005 cette respectabilité retrouvée en plaçant la question "Quelle est la base biologique de la conscience ?" parmi ce qu'il considère être les vingt-cinq plus grandes questions de la science pour le prochain siècle (Miller, 2005).

L'évolution de la pratique scientifique est indubitablement pour quelque chose dans ce retour en grâce : les approches introspectives sont dans l'ensemble passées de mode. Il s'agit désormais de produire des théories reliant différents types de *données publiques*, données réunies autour de l'usage d'une expression : être conscient. Parmi ces

1.1. LE RETOUR EN GRÂCE DES INTERROGATIONS SUR LA CONSCIENCE⁷

données se trouvent d'un côté : les énoncés verbaux des sujets, les performances lors de tâches perceptives et cognitives, et de l'autre : la physiologie de notre organisme et l'organisation fonctionnelle du système nerveux, diverses variables sur l'activité corticale pendant l'exécution de diverses tâches. Derrière cette pratique officielle, cependant, le débat est houleux.

Pour certains, le fait que la conscience soit de nouveau débattue par les scientifiques tient lieu de retour en grâce pour des "états conscients" qui auraient été discrédités à tort par les behavioristes ([Koch and Crick, 2004](#)) :

After a hiatus of half a century or more, the physical origins of consciousness are being, once again, vigorously debated in numerous books and monographs published in the last decade. What is it that we can we ascertain about the neurobiological basis of consciousness, in particular visual consciousness, and what can we reasonably assume at this point in time ? [...] Firstly, and most importantly, that there is something to be explained. That the *subjective content associated with a conscious sensation exists and has its physical basis in the brain*. Philosophers refer to this as qualia, what-it-is-like, the aboutness of consciousness and other terms.

Le système nerveux, pour ceux-là, renferme l'explication d'une chose évasive qu'ils désignent sous le nom de "contenu subjectif de notre conscience". Ils en cherchent la base biologique parmi différentes propriétés particulières de la physiologie du cerveau – les oscillations à 35-75Hz dans le cortex sensoriel ([Crick and Koch, 1990](#)), le caractère récursif et parallèle de l'adaptation de différents groupes neuronaux ([Edelman, 2003](#)), le traitement récursif de l'information sensorielle induit par les connexions des différentes aires corticales ([Lamme and Roelfsema, 2000](#)), ou même les phénomènes quantiques à l'oeuvre dans les micro-tubules des neurones ([Hameroff, 2001](#)) (voir par exemple [Chalmers \(2000\)](#) pour une liste étendue de propositions).

Pour d'autres, si les scientifiques doivent chercher à rendre compte des données publiques, dont font partie en premier lieu les énoncés introspectifs des sujets, ils n'ont pas pour cela à découvrir un phénomène profondément mystérieux attaché au système nerveux. La "conscience" échappe aux instruments physiques non parce qu'elle est une "chose d'une nature subjective", mais parce qu'il ne s'agit pas d'une *chose* : il s'agit d'un ensemble de compétences. Un psychologue étudiant la conscience, dit Daniel Dennett, est un peu comme un anthropologue étudiant la mythologie d'une tribu

indigène : il ne lui appartient pas d'expliquer l'*existence* douteuse d'un dieu Esprit, mais de détailler l'origine et le fonctionnement d'un *système de convictions* bien réel et relativement cohérent en un dieu Esprit (Dennett, 1992).

La physiologie de notre organisme, et tout particulièrement du système nerveux et du cortex, est importante pour comprendre l'ensemble complexe de compétences – que nous classifions comme perceptives, cognitives, motrices – caractérisant notre activité. Mais une fois ces compétences et leurs interactions expliquées, il n'y a pas de mystérieuse propriété du système nerveux à débusquer, pas de lien “psycho-physiologique” générant des “états conscients composés de sensations” à partir de l'activité de notre cortex.

1.2 Que doit expliquer une théorie de l'expérience sensorielle ?

L'étude scientifique de la conscience doit rendre compte, entre autre, des énoncés verbaux des sujets. Le problème est que, si les sujets formulent un discours relativement cohérent, ils n'ont pas pour autant la maîtrise de son sens. Ainsi vont-ils affirmer qu'ils ont l'expérience d'une scène visuelle détaillée (Noë and O'Regan, 2000), qu'ils savent raisonner (Griffin and Kahneman, 2002), ou qu'ils ont un libre arbitre (Wegner, 2003). Il ne fait pas de doute que ces énoncés, formulés en quantité par le scientifique en premier lieu, réfèrent à quelque chose, mais à quoi ?

Pressés de s'expliquer (en réalité, se pressant eux mêmes de s'expliquer, tant ces idées, pour vagues qu'elles sont, jouent un rôle essentiel dans la conception qu'ils ont d'eux mêmes), pressés de s'expliquer, donc, les sujets entrent dans un exercice de conceptualisation d'eux-mêmes pour expliquer leur discours, mais la science n'a pas à accorder une foi aveugle aux *théories* qu'ils avancent alors, surtout lorsqu'elles paraissent contradictoires avec des faits expérimentaux. Les sujets ne sont pas en position privilégiée pour fournir une explication rationnelle de ce qu'ils veulent dire ou de ce qu'ils font, ils ne sont en position privilégiée *que pour dire et pour faire*.

Les sujets produisent des énoncés et des comportements autour de termes comme “conscience” ou “expérience”, et la charge est à la théorie rationnelle et expérimentale de les expliquer, c'est-à-dire de préciser en particulier :

1. à quoi *réfèrent* les termes de leur discours,

1.2. QUE DOIT EXPLIQUER UNE THÉORIE DE L'EXPÉRIENCE SENSORIELLE ?9

2. les *conditions de production* de tel genre d'énoncé dans un contexte donné, plutôt que de tel autre.

* * *

Le terme “conscience” a, dans le langage commun, différents sens plus ou moins vagues (Block, 1995; Rosenthal, 2002). L'un de ces sens renvoie à la disponibilité de certaines informations pour le raisonnement et le contrôle rationnel de l'action, comme dans l'expression “avoir conscience d'être dans une situation dangereuse”. On parle alors de *conscience d'accès*. Un autre sens renvoie à ce que les philosophes nomment *conscience phénoménale*, et les psychophysiciens l'*expérience sensorielle* : la “qualité sensorielle” qui accompagne la conscience d'accès à certaines choses – ce qu'est pour nous, “en tant qu'expérience sensorielle consciente”, le rouge, un fa dièse, ou une rage de dent.

Plusieurs théories ont été développées pour rendre compte des données collectées autour de la conscience d'accès. Les théories de l'exécutif central (Baddeley, 1998) et de l'espace de travail (Baars, 1993; Dehaene and Naccache, 2001; Baars, 2005), en particulier, ont reçu une attention considérable, mais il en existe d'autre¹.

Il y a encore beaucoup d'interrogations et de débats sur la conscience d'accès ; cependant, comme l'écrivent Dehaene and Naccache (2001), il n'y a aucun *mystère*, il s'agit d'un challenge empirique tout à fait comparable à ceux que l'on trouve dans d'autres secteurs de la science. L'expression “conscience d'accès” est synonyme d'un ensemble de *capacités* démontrées par un individu, et rien ne laisse à penser qu'il faille révolutionner notre vision du monde pour préciser scientifiquement l'articulation entre notre structure physiologique et l'ensemble des capacités que nous déployons. En fait, la preuve de l'absence de mystère est que nous pouvons concevoir des dispositifs artificiels capables d'extraire et recouper des informations sur l'environnement, utiliser ces informations dans un processus de raisonnement, et contrôler l'action en fonction de ce processus.

Les énoncés collectés autour du second usage, eux, soulèvent par contre encore des controverses fondamentales au niveau même de leur référent : *de quoi parlent donc les*

¹e.g. Shallice (1988); Posner and Rothbart (1998); Edelman (2003). En réalité, toutes les théories neuroscientifiques de la conscience sont, malgré parfois les affirmations de leurs initiateurs, des théories de la conscience d'accès (Block, 1995).

sujets lorsqu'ils parlent de leurs "sensations"², lorsqu'ils parlent de leurs similarités et de leurs différences ? Quel est l'objet de leur discours, comment expliquer ce discours ? Plusieurs siècles d'études et de données expérimentales n'ont pas permis d'arriver à une forme de consensus sur la question. Dans le domaine, semble-t-il, chaque argument a son contre-argument, et différentes conceptions du problème cherchent à occuper la place sans qu'il paraisse raisonnable ni de les balayer complètement du revers de la main, ni de les prendre tout à fait au sérieux.

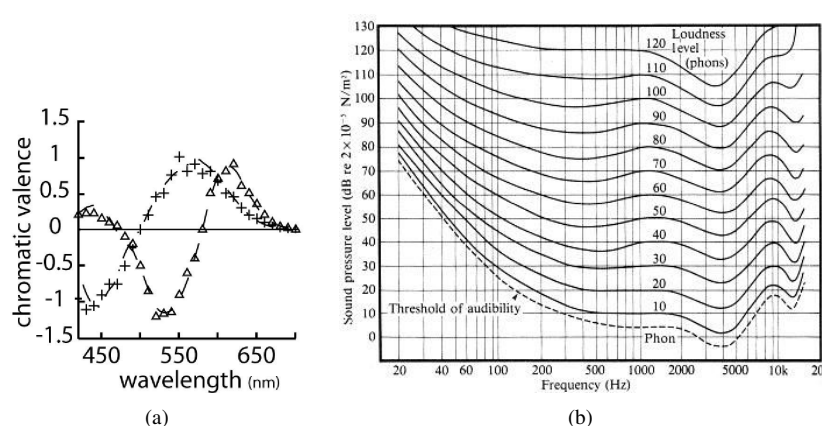


FIG. 1.1 – Deux ensembles typiques d'énoncés quantitatifs sur les sensations. a) Annulation des teintes opposées. Triangles, ordonnées positives : "annulation de la sensation de rouge". Rapport d'intensité nécessaire entre une lumière de 520 nm (verte) et une lumière de la longueur d'onde spécifiée en abscisse pour que leur superposition soit jugée par le sujet comme ne "contenant" plus de rouge (i.e. ne paraisse ni rouge ni verte). Triangles, ordonnées négative : "annulation de la sensation de vert" avec une lumière de 680 nm (rouge). Croix, ordonnées négatives : "annulation de la sensation de bleu" avec une lumière de 580 nm (jaune). Croix, ordonnées positives : "annulation de la sensation de jaune" avec une lumière de 450 nm (bleu). b) Courbes de Fletcher-Munson, indiquant la pression sonore nécessaire à différentes fréquences pour que le son soit jugé par le sujet comme ayant la même intensité.

Les énoncés sur l'expérience sensorielle concernent entre autre la distinction des expériences sensorielles en termes de modalités (e.g. visuelles, auditives, gustatives), l'utilisation de certains qualificatifs pour parler de ces expériences (spatialement étendu, temporel, continu, détaillé, etc), la catégorisation des diverses expériences d'une même modalité (e.g. la catégorisation des couleurs, celle des sons, celle des saveurs), la no-

²Du fait même du débat dont cette partie de ma thèse cherche à rendre compte, il y a dans la littérature un usage ambigu du terme "sensation" : l'usage du sujet qui dit que voir du rouge est une sensation très différente de voir du vert, et l'usage du physiologiste qui dit que la stimulation des cellules ciliées ou des photorécepteurs sont des sensations auditives ou tactiles. J'ai essayé de manière consistante de n'employer le terme que dans son premier usage, et d'utiliser l'expression de "stimulation sensorielle" pour le second.

1.2. QUE DOIT EXPLIQUER UNE THÉORIE DE L'EXPÉRIENCE SENSORIELLE ? 11

tion de “dimensions perceptives” (e.g. les notions de luminance, saturation, teinte, pour les couleurs, ou de hauteur, timbre, puissance pour les sons), et plus généralement la structure des jugements de similarité perceptive.

Un grand nombre de ces énoncés verbaux ont pu être remplacés par les psychophysiciens, au moyen de divers appareillages, en des *énoncés quantitatifs* (Fig. 1.1). C'est donc l'ensemble de ces énoncés, qualitatifs et quantitatifs, que doit expliquer une théorie scientifique de l'expérience sensorielle.

Il y a peu de doutes que les caractéristiques physiologiques de notre corps, de nos systèmes perceptifs et de notre système nerveux, aient un rôle à jouer dans cette explication. En particulier, elles ont un impact direct sur les limitations de nos capacités de discriminations. Il y a en réalité également peu de doutes, bien que cela soit implicitement mis en cause par certaines approches internalistes, que les caractéristiques de notre environnement physique aient également un rôle à jouer (aurions-nous véritablement l'expérience visuelle que nous avons si, tout en étant ce que nous sommes, les lois de la lumière étaient fondamentalement différentes de celles qui régissent notre monde ?). Le véritable débat n'est donc pas tant de discuter la pertinence de telle ou telle pièce isolée du puzzle, faits sur notre physiologie ou sur le monde dans lequel nous évoluons, que sur la manière dont il faut reconstituer le puzzle à partir de ces différentes pièces. Le débat est de proposer un cadre général dans lequel les données empiriques, physiologiques et physiques, deviennent des explications de nos affirmations sur l'expérience.

Chapitre 2

Expérience sensorielle et système nerveux

Le photographe et le chercheur continuent leur promenade. Visiblement insatisfait, le photographe reprend :

Φ Admettons qu'il n'y a pas d'images-vues, et que l'esprit est juste une manière de parler de nos capacités. Mais il y a des informations sur ça [il montre la scène] dans mes neurones, il y a... il y a des *images-encodées* dans ma tête. Elles sont là sous forme d'impulsions électriques [il tapote sa tempe de l'index]. C'est grâce à ça que je vois. C'est ça que tu cherches à expliquer, le grand mystère de la conscience : pourquoi est-ce que je vois des images continues et colorées, alors qu'il n'y a que des impulsions électriques dans mon cerveau ? C'est un phénomène neuro-quantique qui donne la continuité et les couleurs ?

Ψ Pas du tout. Tu ne *vois* pas ce qui se passe dans ton cerveau, tu n'as pas le droit d'utiliser l'expression "je vois X" *à la fois* pour ça [il montre la scène], et pour ce qu'il y a dans ta boîte crânienne [il montre sa tempe]. Je pensais qu'on était d'accord sur ce point : comme tu me montres ça [il fait un geste insistant vers le paysage] pour dire que c'est ce que tu vois, ton discours réfère à ça.

Φ [après une pause]... mon discours parle peut être de ça [il montre le paysage], mais moi, je ne peux parler que de ce qui est *en moi* ! C'est mécanique : je ne peux pas saisir ce qui est devant moi par magie, sans que cela passe par des modifications de mon système nerveux. Ce sont ces modifications que je ressens, en vérité.

Ψ [agacé] Voyons, tu ne peux pas dire autre chose que ce que dit ton discours. Tu dois pouvoir t'engager dans une certaine structure de fonctionnements pour affirmer que tu vois ce paysage ou que tu ressens des choses. Tu dois pouvoir t'engager dans des fonctionnements qui reposent entre autre sur des modifications de ton système nerveux. Mais ça ne veut pas dire que tu parles de ce qui se passe là [il pointe son index sur front].

Φ [sourcils froncés, l'air buté] Mais enfin, ce dont j'ai conscience est bien dans mon cerveau ! Si on stimulait mon cortex visuel, j'aurais une expérience visuelle. Si on stimulait mon cortex auditif, j'aurais une expérience auditive. C'est bien dans mon cerveau qu'est l'explication de l'expérience sensorielle, de ces *qualités* particulières qu'ont les sensationnelles et auditives ! Comment est-ce que tu peux être d'aussi mauvaise foi ?

Ψ Stimuler telle ou telle aire corticale peut être suffisant pour que tu te trouves entretenir la conviction de pouvoir t'engager dans telles ou telles structures de fonctionnements par rapport à ton environnement, par exemple celles qui correspondent à voir un flash ou entendre un son. Mais cela ne signifie pas que tu parles de ce qui se passe dans ces aires quand tu parles de ce que tu ressens.

Φ [désemparé] Mais alors, de quoi est-ce que je parle, comment peut-on expliquer ce que je ressens ?

Ψ A mon avis, il faut chercher à comprendre le sens du discours : à quelles structures de fonctionnements renvoient des expressions comme voir un flash ou entendre un son. Quand tu dis que voir et entendre sont des expériences différentes, je pense que tu parles de différences et de similitudes qu'il est possible d'établir entre ces structures, comme quand tu dis que ça et ça [il montre le paysage et l'écran de l'appareil photographique], c'est similaire.

Φ [il secoue la tête] Enfin, une structure de fonctionnement, ça n'est pas une expérience comme voir du rouge !

Ψ Pas n'importe quelle structure de fonctionnement, bien sûr. C'est une des choses qu'il faut chercher à comprendre : à quel type particulier de structure de fonctionnement nous faisons référence quand nous parlons de notre expérience. Il faut que ce soit une structure qui permette d'expliquer tout ce qu'on affirme à propos des sensations. [il réfléchit un instant] Et entre autre : que tu es convaincu que ce n'est pas un fonctionnement.

Φ [rageur] Mais il n'y a rien à expliquer, c'est comme ça, c'est tout : quand je pense à un fonctionnement, ça n'a rien à voir avec une expérience sensorielle ! Un fonctionnement, c'est juste un fonctionnement, ça n'a pas cette présence, cette réalité concrète, qu'a une expérience sensorielle.

Ψ A mon avis, *penser à un fonctionnement* est justement un fonctionnement très différent du fonctionnement qui est voir du rouge, alors tu dis que c'est différent, et tu as bien raison. [il fait une pose] Oui, vraiment, tu as bien raison. [Après quelques secondes] Tout cela est cohérent avec la théorie que voir n'est rien d'autre qu'une structure de fonctionnement. [il délaisse ostensiblement la conversation, et reprend sa marche en observant distraitemment le paysage.]

2.1 Le siège de l'esprit

La relation entre ces deux sujets de bavardage que sont la physiologie de notre organisme et la conscience semble hanter l'homme depuis qu'il s'étudie. Notre système nerveux jouant un rôle critique dans l'exercice de nos compétences perceptives et rationnelles, comme le révèle les symptômes associés à ses lésions, il y a une rhétorique très ancienne selon laquelle il est "le siège de l'esprit", un lieu géographique de l'esprit : Alcmeon de Crotone fut un des premiers à formuler une affirmation de ce genre six siècles avant notre ère (Gomperz, 1910). En conséquence, la tradition selon laquelle le discours sur les sensations réfère à l'activité elle-même du système nerveux, ou à quelque chose généré par elle, s'est très fermement installée dans la pratique scientifique.

Lorsque les physiologistes ont mis en évidence, à la fin du XIXe siècle, qu'une même stimulation (e.g. une pression mécanique) de nerfs sensoriels différents conduisait les sujets à déclarer qu'ils éprouvaient différentes expériences sensorielles (e.g. la pression du nerf optique conduisant à une sensation visuelle, la pression du nerf auditif à une sensation sonore), ils ont naturellement pensé avoir démontré qu'une expérience sensorielle était l'activation de ces nerfs. Johannes Müller est célèbre pour avoir formulé à l'époque sa théorie de *l'énergie spécifique des nerfs* (Müller, 1826) :

La considération de ces faits ne peut que conduire à la supposition que [...] les nerfs sensoriels ne sont pas de simples conducteurs passifs, mais que chaque nerf sensoriel particulier a un certain pouvoir ou certaines qualités

que son excitation ne font que manifester. Les sensations, par conséquent, consistent [...], pas en la qualité ou état du corps extérieur, mais en la condition des nerfs eux-mêmes, excités par une cause extérieure.

Pour Müller, l'excitation des nerfs sensoriels n'est pas simplement une partie des mécanismes de la production contextuelle de nos énoncés, elle est en réalité *le sujet de ces énoncés* : notre discours affirme que voir et entendre sont des expériences différentes parce qu'il réfère "à la conditions de ces nerfs" et que ces "conditions" sont différentes.

La théorie de l'énergie spécifique des nerfs, bien que considérée comme douteusement spéculative dès sa formulation (Wundt, 1886), a eu une influence importante : les expériences démontrant la ségrégation du traitement de l'information sensorielle ont régulièrement été présentés comme une démonstration de la validité de cette théorie (par exemple, pour des cas particulièrement explicites, dans Adrian (1928), ou plus récemment dans Barlow (1972)). Elle a mené nombre de scientifiques à considérer que l'étude des sensations consistait à contrôler la stimulation physique des sujets, isoler les récepteurs et nerfs réagissant à cette stimulation, puis corrélérer la réponse des récepteurs avec les rapports verbaux des sujets. Avec cette méthode, pensaient ces scientifiques, un nombre fini de sensations seraient isolées, dont toutes les perceptions seraient construites (Reed, 1990).

Ce programme a été un échec. Nous avons considérablement progressé dans notre connaissance des récepteurs sensoriels en tant que systèmes biologiques, mais un siècle de psychophysique n'aura pas permis de se mettre d'accord sur quelles sensations existaient exactement, et associées à quels stimuli et nerfs sensoriels (Watson, 1913; Gibson, 1992).

Du point de vue contemporain, de nombreux exemples ayant montré que la capacité de percevoir même les plus simples attributs perceptifs évolue en fonction de l'expérience (voir par exemple Goldstone (1998) ou Gilbert et al. (2001) pour des revues), il paraît en réalité utopique de s'attendre à une classification établie des sensations, figée et uniforme d'un individu à l'autre, dictée par les seuls récepteurs sensoriels. Le développement d'une expertise individuelle étant de plus influencé par un contexte social et culturel¹, les récepteurs et systèmes sensoriels primaires paraissent susceptibles d'ex-

¹ par exemple la discrimination plus ou moins fine de certains phonèmes dictée par la nécessité de parler une langue donnée (Maye et al., 2002), les interactions entre l'existence de mots pour désigner les stimuli et la catégorisation perceptive (Davidoff, 2001), ou l'influence des associations fonctionnelles évoquées par les odeurs (nourriture, cosmétique, produits d'entretien) sur leur catégorisation perceptive (Chrea et al., 2004)

pliquer certaines contraintes limitatives sur nos capacités discriminatoires, plutôt que d'être l'explication finale de nos affirmations sur les sensations.

En tout état de cause, de nombreuses données sur les sensations ne trouvent aucune sorte d'explication dans l'étude des systèmes sensoriels précorticaux. Ainsi, les données sur un statut sensoriel particulier de certaines couleurs² ne correspondent à aucune particularité de la représentation de l'information chromatique au niveau réceptoral, rétinien, ou thalamique (Valberg, 2001). Ou encore, dans le cas de la musique ou de la parole : la hauteur perçue n'est corrélée à aucun signal identifiable au niveau des nerfs auditifs, ni même au niveau du colliculus inférieur ou des corps genouillés médians (Griffiths et al., 2001).

Enfin, l'idée que l'expérience est une forme de connaissance directe de l'état de notre système sensoriel est d'une manière générale en contradiction avec tout un corpus d'expériences. Par exemple, les expériences sur l'amorçage (Musen and Treisman, 1990; Wiggs and Martin, 1998) et sur certains effets consécutifs (He et al., 1996; He and MacLeod, 2001) ont en effet montré qu'une quantité d'information est présente dans le système sensoriel des sujets, et exploitée, sans qu'ils en aient conscience. Inversement, l'information rétinienne étant limitée par une anatomie pour le moins surprenante et nos yeux bougeant sans cesse (Sec. 2.2, en particulier la figure 2.4), nous affirmons avoir conscience d'informations qui ne peuvent mécaniquement pas être présentes à un niveau précortical.

* * *

Pour ces différentes raisons, la théorie initiale de Müller, focalisée sur les nerfs sensoriels, est aujourd'hui largement considérée comme intenable. Une version remaniée de cette théorie, cependant, continue de sévir dans plusieurs livres standards de neurosciences (e.g. dans Kandel et al. (2000)). Cette version garde l'idée que les sensations sont expliquées par le fait que différentes parties du système nerveux "ont un certain pouvoir ou certaines qualités que leur excitation ne font que manifester". Le remaniement consiste à considérer que ces parties très spéciales du système nerveux sont des aires corticales complexes plutôt que de vulgaires nerfs sensoriels.

On sait en effet depuis longtemps, depuis les travaux de Penfield (1959), que la stimulation électrique directe du cortex peut induire de riches expériences phénoménales.

²Les couleurs rouge, vert, bleu et jaune seraient "pures", et les autres couleurs composées

Les lésions de certaines aires corticales résultent par ailleurs en une altération sélective de l'expérience phénoménale, par exemple : cécité corticale (Weiskrantz et al., 1995), agnosie visuelle (Riddoch and Humphreys, 1987), achromatopsie (Zeki, 1993), akinétopsie (Regan et al., 1992). Enfin, les enregistrements par électrode et les études d'imagerie fonctionnelle ont montré qu'il était possible d'établir des relations univoques entre l'activité de certaines aires corticales (Zeki and Marina, 1998; Näätänen, 2001; Kanwisher and Yovel, 2006), voire de certains neurones (Logothetis, 1998; Rolls et al., 1999; Koch and Segev, 2000; Rees et al., 2002), et la conviction du sujet d'être dans certaines situations perceptives : voir des couleurs, entendre des sons, ou même voir des visages (Fig. 2.1).

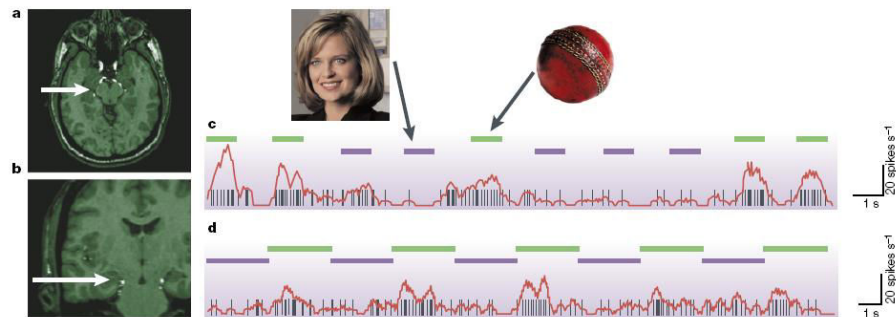


FIG. 2.1 – Corrélation entre l'activité d'une cellule et la conscience de deux scènes visuelles (Rees et al., 2002). a & b) position de l'électrode dans le plan axial et coronal. c) Potentiels d'action du neurone et rythme de décharge lorsque deux scènes perçues sont alternées, d) Potentiels d'action du neurone et rythme de décharge lorsque les deux scènes sont imaginées en alternance.

Dans un certain discours scientifique, le siège de la conscience a en conséquence été déménagé des nerfs sensoriels vers d'autres quartiers du cerveau. Les données sur l'expérience sensorielle devraient donc trouver leur explication dans l'étude de ce qui a été nommé les *Corrélat Neuronaux de la Conscience* (NCCs, Fig. 2.2) : ces mécanismes corticaux dont l'activité est corrélée avec les données, verbales ou comportementales, qualifiant telle et telle information comme consciente (Zeki and Bartels, 1998; Frith et al., 1999; Rees et al., 2002; Crick and Koch, 2003).

Identifier et étudier les NCCs, semble-t-il, est la manière prudente d'aborder la question de l'expérience sensorielle (Churchland, 1994) : en laissant aux philosophes et aux psychologues ce plaisir désuet de ratiociner sur ce qu'elle est ou n'est pas, en avan-

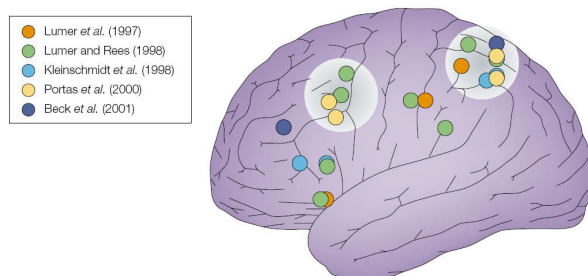


FIG. 2.2 – Corrélats Neuronaux de la Conscience au niveau pariétal et préfrontal (Rees et al., 2002). Aires du cortex dont l’activation est corrélée avec des changements d’expérience visuelle. Un nombre important d’aires sont situées dans le cortex pariétal supérieur et dans le cortex dorsolatéral préfrontal.

çant implacablement dans la compréhension du système biologique que nous sommes, et en pariant sur les découvertes incroyables qui ne manqueront pas de surgir de l’étude de cette “base biologique de la conscience (phénoménale)” (Miller, 2005).

Pourtant, de l’aveu même des plus fervents partisans de l’approche, il n’existe aujourd’hui aucun cadre conceptuel pour formuler une explication des données sur les sensations à partir de l’étude des NCCs (Crick and Koch, 2003) :

The most difficult aspect of consciousness is the so-called ‘hard problem’ of qualia – the redness of red, the painfulness of pain, and so on. No one has produced any plausible explanation as to how the experience of the redness of red could arise from the actions of the brain.

En toute candeur, on peut remarquer que les corrélats neuronaux de la conscience du rouge et du vert, différentes activités neuronales dans V4 (Bartels and Zeki, 2000), ont selon toute vraisemblance *exactement* les mêmes propriétés physiologiques : pourquoi certaines activités devraient “produire” une sensation de rouge et les autres de vert, plutôt que rien, plutôt que l’inverse ? Comment pourrait-on jamais expliquer *ça* ? Il semble que cette approche mène dans un mur, quels que soient les progrès de la science telle qu’on la connaît (Levine, 1983; Nagel, 1974).

L’approche basée sur les NCCs repose en réalité sur une spéculation pour le moins audacieuse (Fig. 2.3) : toutes les données démontrant la corrélations entre certaines activités corticales et la conviction du sujet d’être dans telle et telle situation perceptive montrent uniquement que ces circuits neuronaux jouent un rôle critique dans la capacité du sujet à soutenir certaines convictions vis-à-vis du sujet de son discours, et cela est distinct de l’affirmation selon laquelle le sujet est en train de parler de ce qui se passe

dans ce circuit, sans parler d’une chose “générée” par ce qui se passe dans ce circuit. En un mot, le fonctionnement de ces aires corticales est indéniablement un rouage critique de cet ensemble de capacités que l’on nomme la conscience d’accès, mais cela n’implique pas que les NCCs renferment l’explication de la conscience phénoménale – l’explication des données sur les sensations.

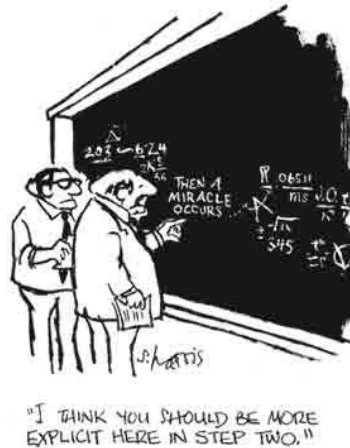


FIG. 2.3 – Le problème avec les NCCs. D’après [Dennett \(1992\)](#)

2.2 Les propriétés “paradoxaes” du système visuel

En réalité, le problème de la théorie de Müller n’était pas l’endroit où ses connaissances l’avaient amené à héberger la conscience. Le problème est en effet le même que l’on pense que la conscience émerge de l’activité des nerfs sensoriels, du cortex pariétal ou dorsolatéral, ou des oscillations à 35-75Hz. C’est le problème de penser que la conscience d’une chose est elle-même “une chose”, plutôt qu’une manière de fonctionner vis-à-vis des choses. Cette approche amène inévitablement à chercher en nous “une chose” qui a les mêmes caractéristiques que ce dont on parle, par exemple d’une image – ce qu’on montre comme étant une image. Cette circularité, comme le montre [O’Regan \(1992\)](#), apparaît de manière quasi-caricaturale dans un ensemble de recherches menées sur la perception visuelle, en particulier sur notre impression d’une scène visuelle continue, détaillée, homogène, stable.

Les recherches en question ont pour point de départ des caractéristiques jugées “paradoxaes” du système visuel, et que Helmholtz avait résumées d’une manière célèbre

en écrivant :

Il n’est pas exagéré de dire que si un opticien voulait me vendre un instrument avec tous ces défauts, je devrais, avec raison, sévèrement reprocher sa négligence et lui retourner aussitôt son instrument.

Parmi ces caractéristiques paradoxales vient en premier lieu la perte significative d’information entre l’image projetée par le cristallin et les signaux délivrés par l’ensemble des photorécepteurs, en raison du dense réseau d’axones et de vaisseaux sanguins qui se trouvent *devant* les photorécepteurs de la rétine (Fig. 2.4(a) et 2.4(c)). Une partie de la rétine, la tache aveugle, est même complètement dépourvue de photorécepteurs parce qu’il y passe les axones de toutes les cellules ganglionnaires, ainsi que les vaisseaux sanguins irriguant la rétine. Une deuxième caractéristique surprenante est la répartition des photorécepteurs, fortement inhomogène (en type et en densité spatiale) au travers de la rétine (Fig. 2.4(b) et 2.4(e)). Une troisième est l’importance des aberrations chromatiques du système oculaire, qui induisent une différence de réfraction de l’ordre de 1.5 dioptries entre les longueurs d’ondes aux extrémités du spectre visible (Fig. 2.4(d)). Une quatrième caractéristique est que nous produisons entre trois et cinq saccades oculaires par seconde (Fig. 2.4(f)), lorsque le temps d’intégration au niveau rétinien est de l’ordre du dixième de seconde.

Toutes ces caractéristiques sont paradoxales parce que, ainsi va le raisonnement, nous devrions dans ces conditions avoir conscience d’obstructions dans notre champ visuel, d’inhomogénéité dans la résolution spatiale de nos capacités discriminatoires, et de flou trois à cinq fois par seconde. Différents mécanismes compensatoires seraient donc nécessaires, en particulier un processus de *suppression visuelle* pendant les saccades (Ross et al., 2001), et un processus de *filling-in* pour interpoler au niveau cortical l’information rétinienne manquante (Ramachandran and Gregory, 1991; Weerd et al., 1995; Weerd, 2006).

Il est difficile de ne pas penser, à l’évocation de ces “paradoxes” et de ces mécanismes “compensatoires”, à l’obsession qui fut celle de Léonard de Vinci à propos de l’inversion de l’image rétinienne. Considérant que l’oeil était une sorte de *camera obscura*, comment, se demandait-il, notre expérience visuelle peut-elle être “à l’endroit” si l’image projetée sur la rétine est à l’envers ? Il faut nécessairement qu’un mécanisme redresse cette image, avait-il conclu, en proposant pour ce faire diverses solutions optiques d’une logique pour le moins troublante.

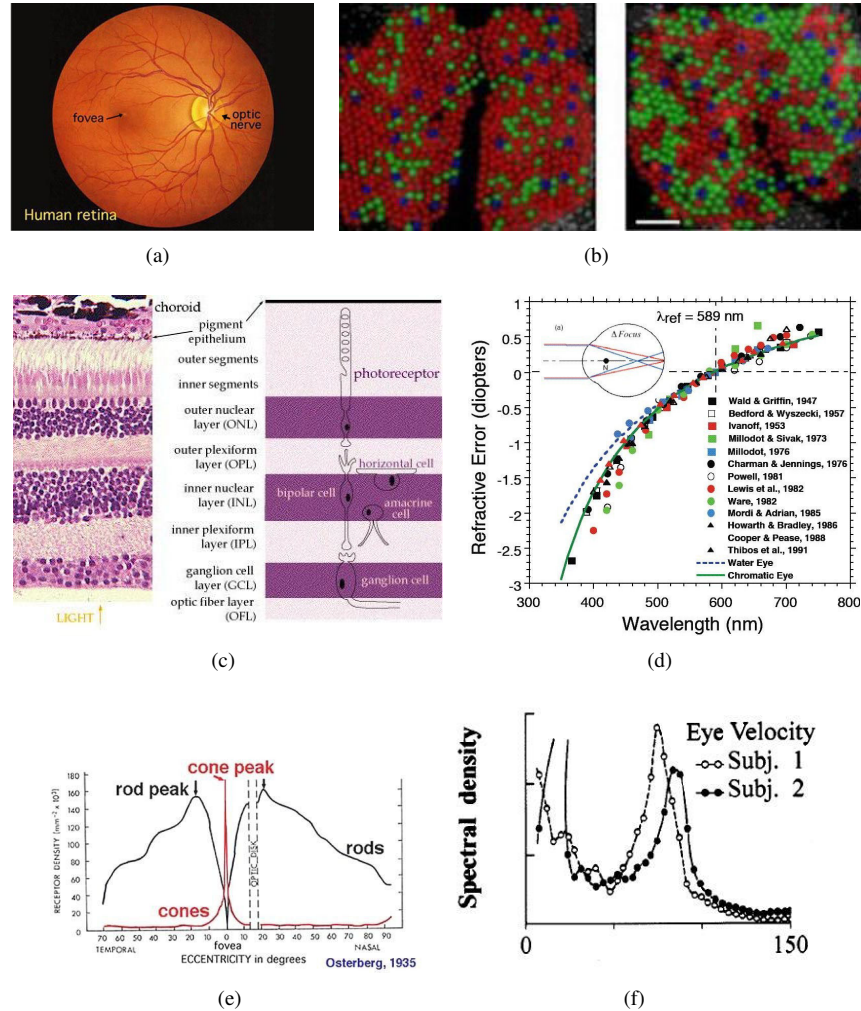


FIG. 2.4 – Quelques unes des caractéristiques “paradoxaes” du système visuel. a) photographie du fond de l’oeil, faisant apparaître les vaisseaux sanguins qui se trouvent devant la rétine. b) répartition fortement inhomogène des types de photorécepteurs chez deux sujets standards, en pseudo-couleurs (rouge : L, vert : M, bleu : S). c) coupe transversale de la rétine faisant apparaître que les photorécepteurs sont tête-bêche par rapport à la provenance de la lumière, et que les différents neurones rétiniens se trouvent *devant* les photorécepteurs par rapport à la lumière. d) aberrations chromatiques : il faudrait une lentille d’environ 2 dioptries pour ramener l’image rétinienne dans le rouge dans le même plan focal que l’image rétinienne dans le bleu. e) densité des bâtonnets et des cônes, faisant apparaître l’absence de bâtonnets dans la fovéa, et l’absence de photorécepteurs au niveau de la tâche aveugle. f) densité fréquentielle des mouvements oculaires (chez le macaque), faisant apparaître deux types de mouvements caractéristiques : des tremblements dus à la tension des muscles oculaires, et les saccades.

L'idée que l'image rétinienne devrait avoir une certaine orientation est la conséquence du raisonnement suivant : lorsque nous avons devant nous un tableau la tête en bas, nous le percevons comme tel, comment dans ces conditions peut-il y avoir une image la tête en bas devant notre rétine sans que nous la percevions comme telle ?

Ce raisonnement est fallacieux parce que, *stricto sensu*, nous ne voyons pas l'image rétinienne : l'image rétinienne n'est une image que pour les scientifiques, parce qu'ils se tiennent (ou s'imaginent) dans une certaine relation par rapport à ce qu'il y a derrière le cristallin d'un sujet. L'image rétinienne n'est pas une image pour le sujet lui-même, qui se trouve dans une relation différente, ne pouvant s'engager dans la même structure de fonctionnements que le scientifique : entre autre, il n'y a pas d'œil dans l'œil pour scanner l'image rétinienne. En effet, les caractéristiques physiologiques du système visuel, en particulier les mouvements de yeux, sont un aspect extrêmement important de ce qu'est voir (Findlay and Gilchrist, 2003; Martinez-Conde et al., 2004), en couplage avec les “traitements” purement neuronaux de l'information : la sélection de l'information visuelle est principalement basée sur ces mouvements oculaires³, qui suivent des règles complexes (Liversedge and Findlay, 2000) et sont étroitement couplés avec, voire constitutifs d'une partie, de nos capacités cognitives (Ballard et al., 1997).

Plus personne ne recherche aujourd'hui un mécanisme de redressement de l'image rétinienne. La communauté scientifique a reconnu qu'il est impossible de poser un cadre conceptuel pertinent pour justifier l'idée du besoin d'un redressement, tout aussi bien que d'expliquer ce en quoi ce redressement pourrait consister au juste⁴. Pas de question, pas d'énigme. Cela ne retire rien à l'intérêt de savoir que l'image rétinienne est inversée : cela explique par exemple qu'une lésion en bas à droite de la rétine induise une incapacité discriminatoire dans le champ visuel supérieur gauche.

Concernant les autres caractéristiques du système visuel, de même, il est crucial de garder à l'esprit que *stricto sensu* nous ne voyons pas la stimulation rétinienne. Notre système visuel n'est pas un périscope pour un autre œil, mais une partie de la machine à qualifier le sujet de notre discours. Il n'y a donc tout simplement pas de sens à s'étonner de nos affirmations selon lesquelles nous ne voyons pas une scène obstruée et mouvante : ce n'est pas l'image rétinienne que nous voyons, et la scène dont

³ même s'il existe également un système de focalisation “virtuel”, capable d'opérer sans mouvement de l'œil, e.g. Posner and Petersen (1990); Desimone and Duncan (1995)

⁴ quoiqu'il soit malheureusement facile de trouver des textes de vulgarisations expliquant que l'erreur de Léonard de Vinci était d'avoir cru que c'était un mécanisme optique qui redresse l'image, alors que serait le cerveau...

nous parlons n'est pas, vis-à-vis de la structure des fonctionnements qui constituent pour nous ce qu'est voir, comme ce que nous désignons être une image obstruée et mouvante.

Il est possible qu'un changement de stimulation rétinienne conduise, dans le contexte d'une saccade, à une absence d'activation au niveau des neurones de MT et MST (Thiele et al., 2002), ainsi que de V4 et V7 (Kleiser et al., 2004), alors qu'une activité systématique est observée dans ces aires lorsqu'un même changement de stimulation est le fait de l'environnement. Il est également possible qu'il existe une représentation rétinitopique de la tache aveugle dans V1, activée par la stimulation du voisinage de la tache aveugle (Fiorani et al., 1992; Komatsu, 2006). De telles caractéristiques contribuent à déterminer la structure de fonctionnements dans laquelle nous sommes engagés vis-à-vis de notre environnement et qui constituent ce qu'est voir. Mais dire qu'il y a là une *compensation* des défauts du système visuel (i.e. une compensation du mouvement des yeux ou de la tâche aveugle), permettant de délivrer une information corrigée au "véritable" organe de la vision dans le cortex, ignore entre autre le fait que le véritable organe de la vision comporte des yeux, mobiles, et un ensemble de mécanismes mis en oeuvre à partir du niveau rétinien : le véritable organe de la vision comporte des yeux, il n'est pas dans notre tête.

* * *

Évidemment, nier les paradoxes précédents, i.e. nier la légitimité d'un certain type d'attentes, ne revient en aucune manière à prétendre que la physiologie et le fonctionnement du système visuel n'ont aucun rapport avec les affirmations que nous tenons sur notre expérience visuelle. Nous qualifions *au moyen* du fonctionnement de notre organisme, et le fonctionnement de la machine à qualifier participe évidemment de manière constitutive, avec l'environnement dans laquelle elle se trouve, à déterminer les capacités de discrimination, de détection, aux "comme si" des situations perceptives. Il est ainsi naturel de s'attendre à ce que l'architecture neuronale de la rétine soit critique pour expliquer certaines confusions perceptives systématiques (Fig. 2.5) – bandes de Mach ou grille de Hermann-Hering (parce que l'inhibition latérale induit certaines équivalences partielles de fonctionnement) –, la non détection de certains stimuli (lorsqu'ils se situent au niveau de la tache aveugle), etc. Les caractéristiques du système sensoriel sont tout à fait fondamentales *par les différences et équivalences de fonctionnement* qu'elles induisent.

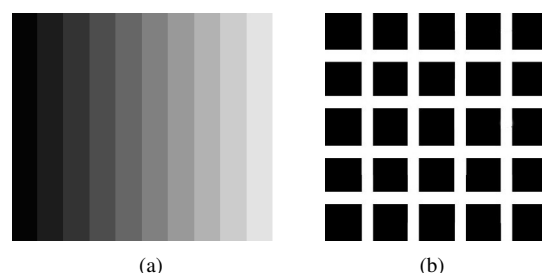


FIG. 2.5 – Bandes de Mach et grille de Hermann-Hering

2.3 L'expérience sans domicile

Deux familles d'expériences éclairent d'une lumière particulière le débat sur la conscience phénoménale. La première famille suggère que la nature visuelle ou auditive de notre expérience est dans une large mesure indépendante des spécificités physiologiques prédéterminées au niveau des différentes aires corticales. La seconde remet en cause l'idée qu'il y a dans notre système visuel une représentation globale du champ visuel et que notre sensation d'un champ de vision unifié et spatialement très étendu réfère à une chose "en nous".

Ces deux classes d'expériences ont suggéré à plusieurs scientifiques que le référent de notre discours sur les sensations est à chercher ailleurs que dans l'état de notre cerveau.

2.3.1 Voir avec le cortex auditif

Pour déterminer s'il existe un lien entre qualité de l'expérience sensorielle et identité du substrat cortical traitant l'information sensorielle, des neurophysiologistes ont réalisé au cours des années 90 une expérience imaginée par William James un siècle plus tôt.

Le célèbre psychologue argumentait que la qualité sensorielle était fonction de l'aire corticale activée, et il avait donc suggéré que si un chirurgien inversait les nerfs optiques et auditifs d'un patient, cette personne verrait le tonnerre et entendrait les éclairs. Alors qu'une telle opération est bien sûr difficile à envisager, on sait cependant que, lorsque les afférents normaux d'un noyau thalamique sont retirés à un stade suffisamment précoce du développement, ce noyau se retrouve innervé par les afférents voisins ([Schneider, 1973](#)). Cette particularité développementale a suggéré la manipu-

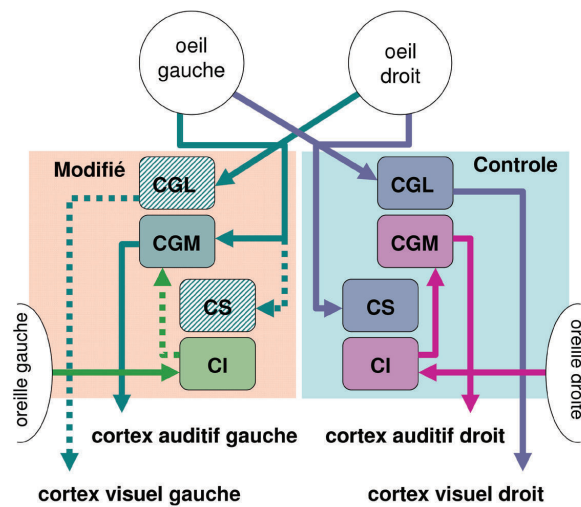


FIG. 2.6 – Manipulation chirurgicale effectuée sur les furets. A droite, situation normale : l'information visuelle est relayée au cortex visuel par le thalamus visuel, l'information auditive en provenance du colliculus inférieur est relayée au cortex auditif par le thalamus auditif. A gauche : la liaison entre CI et CGM est sectionnée par chirurgie, puis le colliculus supérieur est lésionné.

lation suivante (Sur et al., 1999). La liaison entre le colliculus inférieur et le thalamus auditif gauche de jeune furets a été sectionnée par chirurgie, avec pour conséquence de retirer les afférents auditifs à ce noyau thalamique, puis le colliculus supérieur gauche des furets enlevé, ce qui a provoqué l'innervation du thalamus auditif par les afférents du colliculus supérieur. Après cette manipulation, l'information sensorielle visuelle en provenance de l'oeil droit est relayée à la fois aux cortex visuel et auditif gauche. Le côté droit a été laissé quant à lui intact (Fig. 2.6) .

Les furets sont ensuite élevés normalement jusqu'à l'âge adulte, où l'on procède à une lésion du thalamus visuel gauche pour que de ce côté seul le cortex auditif reçoive l'information visuelle. On effectue alors des test comportementaux, en entraînant les furets à aller à droite lorsqu'on leur présente une lumière sur la gauche (information traitée par le cortex visuel droit), et à droite lorsqu'on leur présente un son (traité par le cortex auditif droit). Après cet entraînement, on présente aux furets une lumière à droite (information traitée par le cortex auditif gauche), le postulat étant que les furets généraliseront leur comportement au travers de qualités sensorielles identiques (von Melchner et al., 2000).

Le résultat de ces tests est que les furets vont à gauche, c'est-à-dire que le stimu-

lus lumineux traité par le cortex auditif gauche est considéré par le furet comme un stimulus similaire à celui traité par son cortex visuel droit. Ce qui compte de manière déterminante, semble-t-il, sont les possibilités sensori-motrices que le furet a appris à associer à la stimulation sensorielle, et pas l'identité physiologique des structures corticales impliquées⁵.

2.3.2 La cécité au changement

L'idée que le système visuel construit une carte sensorielle globale du champ visuel, utile pour organiser l'activité cognitive et le comportement, et expliquant notre impression subjective d'une scène étendue et détaillée, est une supposition traditionnelle des scientifiques travaillant sur la vision. David Marr débute ainsi son influent livre *Vision* (Marr, 1982) :

[...] if we are capable of knowing what is where in the world, our brains must somehow be capable of representing this information - in all its profusion of color and form, beauty, motion and detail.

Voir, dit-il, est savoir ce qu'il y a dans le monde, et *savoir est avoir l'information*.

Cette affirmation passée en douce, voir est savoir et savoir est avoir, n'est pas triviale. Dans nombre de situations, le savoir des sujets n'est pas une réponse prête à l'avance qu'ils se garderaient sous la main (ou dans un coin de leur cerveau) dans le seul but de pouvoir la brandir : leur savoir est au contraire la capacité d'élaborer une réponse au fur et à mesure de leurs besoins.

Contre notre intuition la plus profonde, cela semble être le cas de la vision : l'idée qu'il existe en nous une représentation globale et détaillée du champ visuel, résultant de l'intégration de l'information sensorielle obtenue au travers de diverses fixations oculaires (Feldman, 1985), se trouve confrontée à toute une série de résultats contre-intuitifs. Les informations sensorielles ne semblent pas être intégrées au travers des mouvements oculaires (Jonides et al., 1983; O'Regan and Lévy-Schoen, 1983; Irwin, 1991; Rayner and Pollatsek, 1992), et de larges changements dans des scènes naturelles, qui devraient être facilement détectés à partir d'une représentation globale du champ visuel, peuvent passer inaperçus s'ils surviennent pendant une saccade (Grimes,

⁵Cet apprentissage, bien sûr, est réalisé entre autre *au moyen* d'une évolution de l'organisation corticale, et de fait, le cortex auditif des furets présente après ces expériences une organisation typique de celle du cortex visuel (Sharma et al., 2000). Cela est cohérent avec des données classiques indiquant une forte dépendance du développement du cortex visuel vis-à-vis de l'expérience sensorimotrice (Held and Hein, 1963; Hirsch and Spinelli, 1970; Blakemore and Mitchell, 1973; Crair et al., 1998).

1996; McConkie and Currie, 1996; Henderson and Hollingworth, 1999b) ou simultanément à l'apparition de distracteurs (Rensink et al., 1997; O'Regan et al., 1999). Ce fait a été nommé “la cécité au changement” (Fig. 2.3.2).

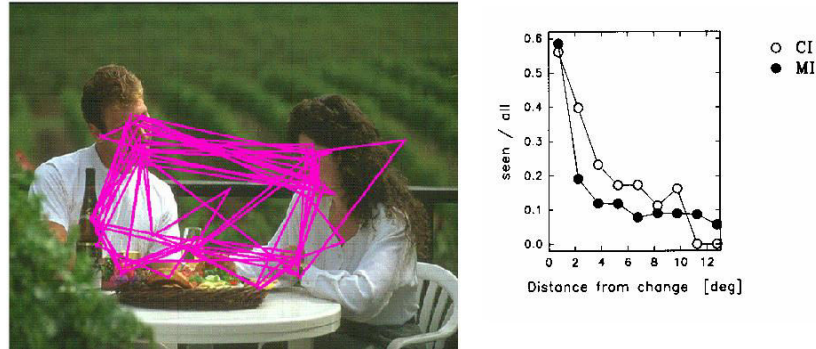


FIG. 2.7 – Cécité au changement (Rensink et al., 1997). A droite, une image test, avec en rose le trajet du point de fixation des yeux. L'expérience consiste à insérer un écran blanc, puis afficher la même image où la barrière du second plan a été déplacée. A gauche, les taux de détection du changement en fonction de sa distance avec le point de fixation.

L'une des explications trouvées pour rendre compte de ces données est que le mode habituel de détection de l'occurrence et du lieu d'un changement repose sur un captage de l'attention par l'activation de détecteurs de transitoires à un bas niveau du système visuel (Kulikowski and Tolhurst, 1973) – plutôt que sur une procédure de comparaison opérant entre des (re)présentations détaillées de la scène visuelle. Les saccades ou les distracteurs perturbent cette procédure, et les sujets doivent alors identifier le changement à partir des informations retenues en mémoire de travail, informations sémantiques dont le schématisme explique le faible taux d'identification observé. Cette interprétation suppose une exploration incessante de la scène qui sert ainsi de “mémoire visuelle externe”. Elle suggère que l'interaction sensorimotrice de notre système sensoriel avec l'environnement est engrené avec le fonctionnement de notre système de planification et de raisonnement de manière à ce que ce fonctionnement soit affecté par les caractéristiques de l'environnement pertinentes à un moment donné pour remplir sa tâche courante (O'Regan, 1992).

Cette explication n'est pas la seule proposition. En fait, il y a un virulent débat sur l'origine de la cécité au changement (Simons, 2000; Simons and Ambinder, 2005) : en particulier, il se pose la question de savoir à quel point l'incapacité de détecter un changement démontre l'absence de représentation de ce qui change – il se pourrait qu'il

existe une représentation corticale très riche, mais qu'il n'existe pas de mécanisme de comparaison entre différents instants. Toutefois, face à l'accumulation de résultats plus ou moins reliés à la cécité au changement, les théories récentes de la perception visuelle ont pour la plupart abandonné l'idée d'une reconstruction détaillée de la scène visuelle, en faveur de celle d'une représentation minimaliste et transitoire limitée à l'objet de l'attention ([Henderson and Hollingworth, 1999a](#); [Rensink, 2002](#)).

Chapitre 3

Retour au réalisme ?

La discussion travaille le photographe, il a l'air profondément perturbé – les sourcils froncés, le regard dans le vide.

Φ Mais enfin, quand je parle des couleurs, par exemple du rouge de cette fleur [il cueille un coquelicot sur le bord du chemin], de quoi est-ce que je parle ? Tu vas me répondre, j'imagine : de ça [il pointe le pétale]. Est-ce que tu veux dire que je perçois les propriétés du monde réel telles qu'elles sont, que je parle des propriétés physiques de ce pétale ?

Ψ [avec un mouvement de négation de la main] Je ne sais pas ce que tu veux dire par “les propriétés du monde réel telles qu'elles sont”. Les physiciens ne sont pas des spectateurs invités dans les coulisses d'un théâtre d'apparences, ce sont des artisans avec des outils, un savoir faire et des capacités d'analogies de plus en plus complexes.

Φ [agacé] Je ne vois pas où tu veux en venir, ça ne répond pas à ma question.

Ψ [continuant l'air de rien] Lorsqu'un physicien dit qu'un objet pèse 3 kg, il dit : tu vas te retrouver dans telle et telle situation si tu mets en oeuvre telle et telle procédure de pesage avec cet objet. Ce système de procédures n'est pas un simple moyen de mesurer la masse, c'est le *référant* du terme masse : pour quelqu'un qui ne sait rien de ces procédures, “cet objet pèse 3 kg” est une phrase vide de sens.

Φ [il commence à s'énerver] Et alors ? Je ne vois toujours pas le rapport ?

Ψ Je pense que lorsque nous disons que ce pétale est rouge [il pointe le pétale] ou que cette tige est droite [il montre la tige], nous parlons des propriétés de notre

environnement, dans la mesure où c'est le même principe de sens du discours que celui des physiciens : je pense que nous exprimons notre conviction de pouvoir nous engager dans des structures de *fonctionnements sensorimoteurs* ayant certaines caractéristiques particulières.

Φ [interloqué] Mais pas du tout. Quand j'ouvre les yeux, je vois une image : c'est étendu, détaillé... Une *image*, ça n'a rien à voir avec une structure de fonctionnements sensorimoteurs !

Ψ Quand tu ouvres les yeux, ta sensation qu'une image étendue et détaillée apparaît vient *justement* du fait que la structure de fonctionnements sensorimoteurs dans laquelle tu peux t'engager évolue presque comme quand ce qu'on appelle une image apparaît. [il fait une pause] Ce que je te raconte est contre-intuitif, mais ça permet d'envisager un sens à des affirmations qui nous semblent élémentaires, et qui seraient extrêmement mystérieuses sans cela. Par exemple : pourquoi nous disons que sentir du mou ou voir du rouge sont des expériences très différentes.

Φ [l'air perdu] ...vraiment ?

Ψ [il plie son bras devant lui] "Sentir le mou", par exemple, renvoie au fait que lorsque tu appuies [il appuie avec son index sur son avant-bras], ça cède d'une certaine manière : il y a tout un ensemble de fonctionnements sensorimoteurs dans lesquelles tu peux t'engager. [il prend le coquelicot des mains du photographe] "Voir du rouge", ça renvoie au fait que lorsque tu bouges, tes stimulations rétiniennes évoluent d'une certaine manière : tu peux t'engager par exemple dans toutes ces fonctionnements sensorimoteurs là [il bouge la tête en fixant le coquelicot des yeux, il déplace la fleur en la suivant du regard, il cligne des yeux]. Les structures des fonctionnements sensorimoteurs "sentir du mou" et "voir du rouge" sont très différentes, nous n'arrivons pas à établir d'analogies entre elles : ce sont des expériences très différentes. Il est possible par contre d'établir certaines analogies entre les structures des fonctionnements sensorimoteurs "voir du rouge" et "voir du vert", alors tu dis que c'est plus similaire.

Φ [l'interrompant] Mais je vois le rouge de ce coquelicot [il pointe un pétale] sans faire ce que tu dis : sans bouger ni les yeux ni le coquelicot. Les sensations, c'est une manière d'être, pas une manière de faire !

Ψ A mon avis, les sensations sont un système d'analogies entre différentes structures de fonctionnements dans lesquelles nous *pouvons* nous engager. Mais ce n'est

pas nécessaire de s’y engager : certaines activités neuronales dépendent des propriétés de l’environnement, et interviennent dans les mécanismes qui te permettent de contrôler et planifier tes actions, d’avoir certaines attentes sur l’évolution des stimulations sensorielles : tu as la conviction pratique que tu *dois pouvoir* t’engager dans tels et tels fonctionnements sensorimoteurs.

Φ [Il se gratte la tête d’un air perplexe] Vraiment, j’ai beaucoup de mal à comprendre ce que tu racontes. Je ne comprends même pas de quoi tu parles avec tes histoires de... fonctionnements sensorimoteurs.

Ψ [la main sur la nuque] Oui, il faudrait préciser la notion de structure de fonctionnements sensorimoteurs. C’est sur ça que je travaille. Ce que je te raconte est l’esquisse d’une direction de recherche, pas une solution clé en main.

Φ [il fait non de la tête, une grimace sur les lèvres] Vraiment, c’est pas clair. Et puis de toute façon, ta théorie est uniquement relationnelle. A la limite, elle pourrait expliquer des comparaisons entre différentes expériences, mais pas ce qu’*est* une expérience considérée toute seule, ce que c’est de voir du rouge, sans référence à rien. Et en fait, c’est ça, le vrai mystère de la conscience !

Ψ [las] Les choses ne ressemblent à rien, sans référence à rien. C’est un jeu de miroir : nous essayons d’utiliser une situation pour en penser une autre. [il appuie sur ses paupières] La relativité est une idée très difficile, l’histoire est jonchée de théories absolutistes : à chaque fois, on a voulu se positionner en dehors de notre condition humaine pour concevoir le problème, en imaginant qu’il existait des références et une ontologie qui la dépassaient, et ça nous bloquait dans des absurdités. Nous ne pouvons pas concevoir ce que c’est d’être en dehors du jeu que nous jouons. Je pense qu’il faut avoir l’humilité d’accepter ça.

Φ C’est vraiment contre-intuitif, ton histoire. [avec un ton de reproche] Ce serait plus crédible si tu expliquais comment le cerveau génère des images-vues.

Ψ [il soupire, et reprend sa marche en silence en se rongant les ongles]

3.1 La phénoménologie de la conduite

L’approche sensorimotrice propose de concevoir l’expérience sur l’exemple de ce que nous considérons être la phénoménologie de conduire une voiture, plutôt que sur ce que nous considérons être la phénoménologie de la vision (O’Regan and Noë, 2001) :

Scientists and philosophers frequently get the phenomenology of experience wrong ; they misdescribe what perceptual experience is like. Consider, as an example, the experience of driving a Porsche and its distinctive qualitative character. What does this feeling consist of ? What is it like ? Notice that, in one sense, there is no feeling of Porsche-driving. That is, the character of Porsche-driving does not consist in the occurrence of a special sort of momentary flutter or bodily sensation.

What defines the character of driving a Porsche, rather, is something more complex. There are characteristic ways in which the vehicle accelerates in response to pressure on the gas pedal. There are definite features of the way the car handles turns, how smoothly one can change gears, and so on. What it is like to drive a Porsche is constituted by all these sensorimotor contingencies and by one's skillful mastery of them, one's confident knowledge of how the car will respond to manipulations of its instruments.

In one sense, then, there is no Porsche-driving quale. After all, what it is like to drive a Porsche depends on these various activities. In another sense, however, one can speak of the qualitative character of driving a Porsche, but this must be understood not in terms of the occurrence of a sensation-like quale in the mind, but rather, in terms of one's comfortable exercise of one's knowledge of the sensorimotor contingencies governing the behavior of the car.

We propose that the same account can be extended to such prototypical visual qualia as "the raw feel of a shade of red".

La sensation d'un conducteur de conduire une voiture ultra-puissante, ainsi va le raisonnement, est son installation dans la maîtrise du comportement de la voiture et sa capacité de le comparer au comportement d'autres voitures : chaque aspect de l'expérience d'un conducteur correspond à des caractéristiques objectives et quantifiables de la manière dont la voiture répond aux commandes¹.

Bien sûr, circonscrire et expliquer un ensemble de mécanismes constituant la capacité du conducteur de maîtriser le comportement d'une voiture – la capacité d'engager le système dans un éventail particulier de dynamiques, d'avoir des attentes sur leur dé-

¹Il faut préciser, me semble-t-il, afin d'éviter un malentendu profond, que le cas du conducteur est un *exemple*, et pas une métaphore : l'exemple concret d'un individu dans un certain environnement constitué par une voiture, et pas la métaphore d'un conducteur-cerveau aux commandes d'une voiture-corps. Dans tout ce qui suit, il est important de garder à l'esprit que les mots doivent être pris au sens littéral.

roulement, de les combiner pour atteindre des buts – est une question difficile. Mais il s’agit d’une question empirique, de ce genre de difficulté dont les physiologistes ont l’habitude, il n’y a pas là de mystère psychophysique.

De même, circonscrire et expliquer un ensemble de mécanismes constituant la capacité du conducteur de comparer différents comportements de voitures – identifier le fait qu’appuyer fortement sur l’accélérateur d’une Volkswagen est similaire à appuyer faiblement sur l’accélérateur d’une Porsche, que tourner le volant de la première demande plus d’effort que la seconde, etc – est difficile mais pas *a priori* insurmontable. Cela peut sans doute s’expliquer à partir de mécanismes d’association et de découplage des différentes choses que l’on peut faire.

Ainsi, continue le raisonnement, dans le cas de la conduite, il n’y a ni oeil intérieur ni phénomène psychophysique, et pourtant il y a un véritable sens pour le conducteur à parler de la sensation de conduire une Porsche plutôt qu’une Volkswagen, et il y a la possibilité d’une explication scientifique du sujet de ce discours : une explication de la sensation du conducteur.

* * *

L’exemple de la conduite suggère de prendre le point de vue que percevoir correspond à une interaction sensorimotrice avec notre environnement, et qu’au coeur de la différence entre les sensations se trouvent différentes structures de *dépendances sensorimotrices* : différents types de modifications de notre stimulation sensorielle en réponse à nos commandes motrices.

Lorsque nous bougeons les yeux, par exemple, l’image projetée par le cristallin sur la rétine se déplace et se déforme, et la stimulation rétinienne se modifie d’une manière particulière, déterminée entre autre par l’amplitude du mouvement, l’optique oculaire, les vaisseaux irriguant la rétine, et la répartition en type et en densité spatiale des photorécepteurs. Lorsque nous bougeons la tête, l’amplitude relative des vibrations sonores qui atteignent nos deux oreilles et leur synchronicité changent, ainsi que leur composition spectrale du fait des pavillons auriculaires : les stimulations cochléaires se modifient d’une manière déterminée par l’amplitude du mouvement, la forme des pavillons, et la position des cellules ciliées le long de la cochlée.

Selon l’approche sensorimotrice, les sensations de voir et d’entendre sont notre installation dans la maîtrise de ces dépendances : la capacité d’en faire usage et de les comparer.

Une telle manière d’aborder le problème offrirait, outre une explication des modalités sensorielles qui n’en appelle pas à un “certain pouvoir ou certaines qualités des nerfs”, une réponse à deux questions élémentaires sur les sensations : pourquoi distinguons nous ces manières de fonctionner que sont penser et percevoir, en affirmant qu’il y a “une qualité” particulière au fait de percevoir, une “présence sensorielle”, et pourquoi affirmons nous que les sensations sont incommunicables (O’Regan et al., 2001) ?

Penser ou se souvenir sont des structures de fonctionnements qui ne sont pas affectés par les mouvements du corps. A l’inverse, parce que les dépendances sensorimotrices dans lesquelles nous pouvons nous engager changent avec la position et la posture du corps, l’expérience sensorielle a ce que l’on peut appeler une *corporalité*. D’autre part, nos systèmes sensoriels ont la capacité très particulière d’interférer sur le déroulement de nos processus cognitifs. Un flash brillant ou un son puissant vont automatiquement déclencher un déplacement de notre attention : nous sommes “connectés” de manière très fiable à certains événements de notre environnement. Cette *capacité d’alerte* n’existe pas dans la structure de fonctionnements à laquelle renvoie le verbe “se souvenir” : nous ne réalisons la disparition d’un souvenir que lorsque nous cherchons à y faire appel. L’approche sensorimotrice suggère ainsi que la “présence sensorielle” n’est rien d’autre que la corporalité inhérente à ce que sont les sensations, et la capacité d’alerte déterminée par la physiologie de notre organisme.

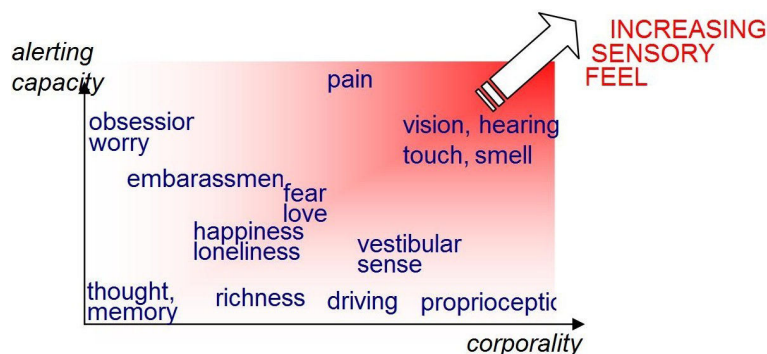


FIG. 3.1 – Corporalité et capacité d’alerte de différentes manières de fonctionner (O’Regan et al., 2001).

La maîtrise de certaines caractéristiques objectives des dépendances sensorimotrices est une maîtrise *pratique*, implicite. Pour reprendre l’exemple du conducteur, ce dernier a accès au fait qu’il peut s’engager dans l’exercice de certaines compé-

tences : tourner, accélérer, changer les vitesses. Il a accès au fait qu'accélérer avec une Porsche ou une Volkswagen sont similaires *modulo* une adaptation de la manière dont il appuie sur l'accélérateur. Toutes ces choses, il peut les faire et il sait qu'il peut les faire, mais il ne sait pas comment il les fait, ni ce que cela signifie physiologiquement de les faire. Il ne peut pas vraiment communiquer "ce que c'est de conduire une Porsche" : tous ses discours, aussi judicieux qu'ils soient, restent différents du fait de savoir *faire* ces choses – dire et faire sont deux fonctionnements différents. Son expérience est incommunicable non parce qu'il chercherait à pointer un monde intérieur que personne d'autre ne peut voir, mais parce que parler d'une compétence sensorimotrice et la maîtriser (avoir développé la capacité de s'engager dans certaines dynamiques sensorimotrice, avec une coordination pertinente de ces dynamiques avec d'autres fonctionnements constitutifs notamment de la conscience d'accès) sont deux choses irrémédiablement distinctes.

3.2 Toile de fond de l'approche sensorimotrice

L'approche sensorimotrice n'est pas isolée dans le paysage historique et contemporain. Poincaré, et la plupart des empiristes avant lui (C.f. section 5.2), défendaient l'idée que la sensation de spatialité réside dans certaines caractéristiques de nos interactions sensorimotrices. Ryle (1949), présentait une conception de ce qu'est voir dans un cadre très similaire à celui qui vient d'être présenté. Merleau-Ponty (1945), cherchait à remplacer l'idée de l'expérience visuelle comme un événement interne, en utilisant des métaphores telles que la palpation de l'environnement. MacKay (1967), en particulier, suggère que la sensation de voir une bouteille est similaire à la sensation d'avoir une bouteille entre les mains : le fait d'être "tuned to" certaines potentialités de notre interaction sensorimotrice, lorsque nous avons en réalité très peu d'informations tactiles à un instant donné.

D'une manière plus générale, l'approche sensorimotrice s'appuie sur une conception active de la perception pour proposer une solution à la question de l'expérience phénoménale. Cela est en phase avec une grande variété de travaux défendant l'importance d'une telle conception pour appréhender le problème de la perception et de la cognition, humaine et animale. Il est impossible de proposer une liste représentative de ces travaux, qui s'inscrivent dans des domaines tels que la psychologie expérimen-

tale de la vision (Findlay and Gilchrist, 2003), la physiologie (Berthoz, 1997; Paillard, 1999), les neurosciences (Sur et al., 1999; Ahissar and Arieli, 2001), les théories de la perception (Gibson, 1979; Beer, 2000), la philosophie de l'esprit (Clark, 1999; Hurley and Noë, 2003; Noë, 2005). Les approches actives de la perception ont également eu un impact important en vision artificielle (Aloimonos et al., 1988; Ballard, 1991) et en robotique (Brooks and Stein, 1994), et leur influence demeure même si les attentes qui ont pu être fondées sur elles ont été reconsidérées sur un pied plus modeste.

L'idée que les sensations dépendent des règles de modifications de la stimulation sensorielle est d'autre part soutenu par des phénomènes tels que la vision faciale, les interactions entre différentes modalités sensorielles, et par les expériences réalisées avec les dispositifs dits "de substitution sensorielle".

La vision faciale est la capacité de certains aveugles à localiser des objets devant eux, au travers de l'impression d'être effleuré sur le front, les joues, et parfois la poitrine (James, 1890). Alors que la vision faciale a longtemps été considérée comme une forme de perception tactile, cette capacité disparaît lorsque les oreilles sont bouchées, et donc les informations sensorielles exploitées sont en réalité délivrées par les stimulations cochléaires (Dallenbach et al., 1944). La classification de cette expérience comme tactile par les aveugles est cohérente avec le fait que le type de dépendances sensorimotrices impliqué est similaire à celui qui constitue la perception tactile. En particulier, ce type de perception repose sur des changements de stimulation tout ou rien en réponse aux mouvements de la tête dans l'axe de l'objet.

La vision est connue pour altérer de manière significative la perception dans les autres modalités : perception de la parole affecté par la vision (effet McGurk) (McGurk and MacDonald, 1976; Munhall et al., 1996), localisation sonore affectée par la vision (l'effet ventriloque) (Alais and Burr, 2004), altération de la localisation proprioceptive et tactile (Pavani et al., 2000). De manière réciproque, l'audition peut affecter la vision (Fendrich and Corballis, 2001; Shimojo and Shams, 2001), ou le toucher la vision (Mascaluso et al., 2000). Ces effets perceptifs coïncident avec le fait que le système nerveux, au travers en particulier du colliculus supérieur (Meredith et al., 1987) et du cortex associatif (Wallace and Stein, 1994), intègre un ensemble d'informations en provenance des différents organes sensoriels. Tous ces résultats, qui montrent le rôle extrêmement prégnant de l'apprentissage de certaines dépendances, suggèrent que les modalités sensorielles, dans le contexte d'une exploitation de l'information sensorielle aussi globale

que possible par le système nerveux, peuvent être plus naturellement conçues comme le résultat d'un mécanisme de distinctions établies *a posteriori* que comme la conséquence automatique de différences physiologiques rigides et préétablies.

Les dispositifs de substitution sensorielle, quant à eux, sont des dispositifs visant à fournir, au travers d'un type de capteurs sensoriels, les informations délivrées par les capteurs d'un autre type. Le TVSS (Tactile Vision Substitution System) est ainsi un dispositif conçu pour les aveugles dans les années 70, qui convertit les images captées par une caméra en des "images tactiles" produite par une matrice 20×20 de vibreurs d'un millimètre de diamètre ([Bach-y Rita et al., 1969](#); [Epstein et al., 1986](#); [Auvray et al., 2007](#)). Selon les études menées avec ce dispositif, les sujets sont capables très rapidement, lorsqu'ils ont le contrôle actif de la caméra, de détecter des cibles simples et de s'orienter vers elles, et de reconnaître des objets usuels. De manière remarquable, cette capacité de reconnaissance des objets s'accompagne d'une mise en extériorité de la perception : l'utilisateur finit par oublier l'aspect tactile des stimulations et rapporte percevoir les objets "devant lui". Par ailleurs, les sujets aveugles découvrent d'eux-mêmes des concepts perceptifs nouveau pour eux, tels que la parallaxe, les ombres, l'interposition des objets ; certaines illusions visuelles classiques sont même spontanément reproduites ([Renier et al., 2005](#)).

Chapitre 4

Perspectives sur les dialogues

J’ai essayé de donner une idée de la difficulté du débat sur l’expérience au travers de scénettes impliquant un chercheur et un photographe. Il s’agit, je l’espère, d’une manière vivante de se plonger dans les méandre d’un problème complexe. En outre, la forme de la scénette repose sur une mise en situation concrète, qui permet de trancher avec un certain type de discours abstrait sur l’expérience, cherchant sa cohérence en vase clos alors que son sens vient de sa référence à ce que l’on peut faire. Dans la phrase : “le photographe dit au chercheur : je vois cet objet [il montre une pierre]”, la narration rend apparent le fait que le problème n’est pas une ontologie de ce qu’est l’expérience, ou une image, ou des objets. Le problème est le fonctionnement, en situation, d’un système discursif et référentiel.

Cette question de rhétorique mise à part, les scénettes évoquent plusieurs débats philosophiques très classiques, qui sont ici pointés très sommairement. L’objectif est de montrer que les répliques ne prétendent bien sûr pas donner un mot final, mais qu’elles proposent plutôt un itinéraire dans la multitude de questions et théories générées par l’idée de l’expérience sensorielle.

L’expérience visuelle, c’est l’image [...] qui m’apparaît quand j’ouvre les yeux

Cette description de l’expérience n’est pas qu’une caricature naïve – un article récent sur les NCCs commence ainsi ([Lamme, 2004](#)) :

As soon as we open our eyes we have a rich experience of the scene that is in front of us. It is as if a picture of the outside world is generated in our

head. Where is this picture coming from ? How do nerve cells generate such an experience ?

La réponse du chercheur est que les deux premières phrases posent des constats qui demandent explication, mais que les deux questions suivantes sont par contre inconsistentes, puisqu'elles oublient le "as if" : il n'y a pas à expliquer la localisation et la génération d'une "comme-si-il-y-avait-une-image".

Ce débat est un raccourci simpliste de l'histoire des *sense-data* (Price, 1950; Russell, 1912). Selon Russell, parce que l'apparence des choses (e.g. la couleur et la forme d'une table) change lorsque nous bougeons, nous n'avons pas connaissance des propriétés du monde par la perception, mais connaissance *d'autre chose*¹ (Russell, 1912) :

Let us give the name of 'sense-data' to the things that are immediately known in sensation : such things as colours, sounds, smells, hardnesses, roughnesses, and so on. [...] It is plain that if we are to know anything about the table, it must be by means of the sense-data – brown colour, oblong shape, smoothness, etc. – which we associate with the table ; but, for the reasons which have been given, we cannot say that the table is the sense-data, or even that the sense-data are directly properties of the table.

La théorie des *sense-data* a été dans l'ensemble abandonnée par les philosophes, entre autre parce qu'elle semble reposer sur le prémisses que la perception est nécessairement véridique. C'est un fait qu'il nous arrive d'être convaincu qu'une table est verte, alors qu'elle paraît bleu après examen, mais découle-t-il véritablement de cela que le véritable objet de la perception est une chose verte immatérielle (Austin, 1962) ? Nous *croyons* voir une table verte, et nous nous trompons.

Plus profondément, il me semble, l'argumentation de Russell paraît reposer de manière fondamentale sur l'incompatibilité entre l'idée que ce qu'on nomme "les propriétés physiques" sont des choses indépendantes de nous, et le fait que ce qu'on perçoit dépend de nous (entre autre, de notre position). Cette conception absolutiste de l'objectivité des propriétés physiques me semble questionnable (c.f. la remarque suivante sur le réalisme), et donc l'incompatibilité qui en découle également.

¹La réflexion de Russell est menée par la volonté d'articuler une distinction entre connaître une chose "de seconde main" (connaissance par description et raisonnement) et la connaître de "première main" (connaissance directe, "par nos sens"). La perception, argumente-t-il, est "une prise de conscience directe, non inférentielle", sans expliciter beaucoup plus avant la notion de prise de conscience directe autrement que par cette opposition avec l'inférence logique. L'idée de connaissance directe ("knowledge by acquaintance"), sera très influente, entre autre sur James (James, 1890).

L'expérience visuelle, c'est l'image qui m'apparaît, à moi - l'image qui apparaît... dans mon esprit. Cette phrase est bien sûr une référence au dualisme : l'idée que l'esprit est un lieu, une substance, une entité, d'une *nature* mystérieuse et incompréhensible. D'un point de vue plus moderne, il me semble, il s'agit de l'affirmation que les *métaphores* (les manières de parler) du lieu, de la substance, de l'entité, peuvent s'appliquer à l'esprit : l'idée que la structure des possibilités concrètes évoquées par ces mots peuvent être transposées d'une manière adéquate pour parler de l'esprit. Inutile de préciser que le dualisme n'est plus une théorie très en vogue dans la communauté scientifique (Dennett, 1992).

La réponse du chercheur est l'attaque initialement menée dans Ryle (1949), dans la droite ligne des idées de Wittgenstein, qui suggère que le terme "esprit" n'a d'autre utilité dans notre discours que de décrire le comportement de certains organismes complexes en termes de stratégie, de capacité d'abstraction, et d'attitudes psychologiques comme la curiosité ou croyance.

L'expérience sensorielle, c'est l'image-vue que j'ai en plus de n'importe quel automate. Il s'agit d'une évocation des discussions, dont les philosophes de l'esprit sont friands, sur le thème des "zombies" (Moody, 1994; Chalmers, 1995a). L'une des manières de convaincre d'un mystère de la conscience, en effet, est d'en appeler à notre capacité d'imaginer un automate que la psychologie expérimentale ne pourrait distinguer de nous (i.e. qu'il démontrerait exactement les mêmes capacités que nous, et tiendrait exactement le même discours que nous), mais qui en réalité ne ferait que *sembler* d'avoir une expérience sensorielle, mais n'en aurait pas – un zombie. Cette créature hypothétique est en réalité une caricature de cas connus de sujets affirmant de manière cohérente ne plus avoir d'expérience phénoménale, tout en démontrant certaines capacités de traitement de l'information sensorielle et de son utilisation pour le contrôle de l'action : on parle de vision aveugle (Braddick et al., 1994).

Du point de vue philosophie, le cas des zombies est censé démontrer que toute explication fonctionnelle de l'esprit est vouée à l'échec. Sa force de conviction, sans doute, repose sur le lien culturel très profond établi entre les catégories du vivant et de l'esprit, de l'inanimé et de l'absence d'esprit. Du point de vue scientifique, les cas de vision aveugle démontrent simplement qu'une partie seulement des capacités de ces sujets sont affectés : en particulier, sans doute, la capacité d'alerte est chez eux

fortement affectée, ainsi également que certaines capacités associatives.

le grand mystère de la conscience : pourquoi est-ce que je vois des images continues et colorées, alors qu’il n’y a que des impulsions électriques dans mon cerveau ? La feinte surprise, à la découverte qu’il n’y a dans notre cerveau rien qui ressemble à ce que l’on perçoit, est une manière classique d’imposer l’idée d’un mystère fondamental de l’expérience, d’un “hard problem” (Chalmers, 1995b). Cette surprise douteuse a déjà été largement évoquée en section 2.

Bien sûr, personne ne s’attend plus, comme c’était le cas d’Empedocles, à trouver dans un crâne humain *exactement* ce que l’on perçoit quand on observe le monde. Mais il semble naturel de s’attendre à percevoir quelque chose qui est *un peu* comme ce que l’on perçoit quand on observe le monde, qui est *structurellement* comme ce que l’on perçoit. Le désir d’expliquer l’expérience par un “isomorphisme psychophysique” est profond, la notion a été, entre autre, défendue en son temps par les tenants de la Gestalt (Köhler, 1938) :

Continuity is a structural trait of the visual field. It is also a structural fact that in this field circumscribed particular percepts are segregated as patches, figures, and things. In both characteristics, we have found, the macroscopic aspect of cortical processes resembles visual experience. To this extent, therefore, vision and its cortical correlate are isomorphic.

En fait, il n’est pas rare de trouver encore, dans des articles contemporains, ce genre d’argumentation (Rees et al., 2002) :

Activity in retinal ganglion cells is necessary for normal vision. But most observers would agree that the receptive field properties of these cells are too different from the properties of visual perception for them to give rise directly to conscious vision.

Le problème, suggère l’approche sensorimotrice, n’est pas de trouver des processus neuronaux tels qu’il existe un isomorphisme entre certaines de leurs propriétés et “les propriétés de notre expérience”, mais de trouver par quels processus nous pouvons identifier un isomorphisme, une similarité structurelle, entre tout ce que nous désignons comme ayant la même propriété, telle que la continuité. Et cela n’implique pas trivialement la nécessité de mécanismes physiologiques ayant cette propriété à un quelconque niveau.

Voyons, tu ne peux pas dire autre chose que ce que dit ton discours. Cette réplique évoque l'argument de Wittgenstein contre l'idée d'un langage privé, l'argument du "scarabé dans la boîte" ([Wittgenstein, 1953](#)) :

293. [...] Eh bien, tout le monde vient me dire qu'il ne sait qu'à partir de son propre cas ce qu'est la douleur ! – Supposons que chacun possède une boîte contenant ce que nous appellerons un «scarabée». Personne ne pourrait jamais regarder dans la boîte des autres ; et chacun dirait qu'il ne sait ce qu'est un scarabée que parce qu'il a regardé le sien. – En ce cas, il se pourrait bien que nous ayons chacun, dans notre boîte, une chose différente. On pourrait même imaginer que la chose en question changerait sans cesse. – Mais qu'en serait-il si le mot «scarabée» avait néanmoins un usage chez ces gens-là ? – Cet usage ne consisterait pas à désigner une chose. La chose dans la boîte ne fait absolument pas partie du jeu de langage, pas même comme un quelque chose : car la boîte pourrait aussi bien être vide. – Non, cette chose dans la boîte peut être entièrement «supprimée» ; quelle qu'elle soit, elle s'annule.[...]

Si on stimulait directement mon cortex, je percevrais des choses qui n'existent pas vraiment. // Tu ne les percevrais pas : tu *croirais* les percevoir, et tu te tromperais.
C.f. la remarque sur les sense-data.

Enfin, une structure de fonctionnement, ça n'est pas une expérience comme voir du rouge ! Cette réplique a pour objectif de représenter dans le dialogue les objections de nombreux philosophes ([Nagel, 1974](#); [Levine, 1983](#); [Block, 1995, 1996](#); [Chalmers, 1995a](#)), selon lesquels le fonctionnalisme est une impasse pour approcher la question de l'expérience sensorielle. La réponse du chercheur est la réponse de la psychologie à la philosophie : les théories philosophiques sont des constructions humaines, soumises aux mécanismes de raisonnement observables chez les sujets. Le discours philosophique, au même titre d'ailleurs que le discours scientifique, est soumis à des règles particulières, et est ainsi sensé avoir une cohérence plus forte que le discours de l'homme de la rue. Mais il n'atteint pas pour autant une vérité finale comme on atteint l'Amérique. Certains philosophes sont convaincus que le fonctionnalisme est une impasse ? Le fait qu'ils aient cette conviction est un fait de plus à expliquer, c'est tout.

Quand je pense à un fonctionnement, ça n’a rien à voir avec une expérience sensorielle ! Cette remarque est inspirée du cas tragique de Mary the scientist (Jackson, 1986). Pour défendre l’idée qu’il y a plus dans l’expérience sensorielle que la connaissance, Jackson évoque le cas d’une scientifique prisonnière d’un monde entièrement gris, mais qui aurait appris tout ce qu’il y a à savoir scientifiquement sur les couleurs. Il argumente que cette scientifique découvrirait quelque chose en étant transférée dans notre monde coloré, démontrant ainsi qu’il y a plus dans l’expérience des couleurs que le savoir de tout ce qu’il y a à savoir sur les couleurs.

Nul doute que Mary découvrirait quelque chose après son transfert : elle découvrirait qu’il n’y a pas que les fonctionnements qui sont *penser* aux fonctionnements qui sont voir les couleurs, mais également les fonctionnements qui sont *voir* les couleurs. Cela ne signifie pas, me semble-t-il, que ce n’est pas une question de savoir.

Les physiciens utilisent une rhétorique très définitive pour parler du monde [...]
Ce ne sont pas des spectateurs invités dans les coulisses d’un théâtre d’illusions, ce sont des artisans L’objectivité de la science est une certaine contrainte de cohérence entre ce que disent les scientifiques et ce qu’ils peuvent faire, ce n’est pas l’accès à “une réalité finale” (c.f. la remarque sur les théories philosophiques).

Ma motivation pour s’engager sur cette digression est l’idée que la phénoménologie naît d’une mécompréhension de ce qu’on appelle la réalité, et réciproquement. L’idée d’une énigme de l’expérience se construit généralement, comme l’a fait entre autre Russell, sur l’opposition entre “un monde réel” et “un monde perçu” (C.f. la remarque sur les sense-data). Les attaques contre la cohérence de cette énigme portent généralement sur le monde perçu : il n’y a pas de sense-data. Or il n’y a pas non plus de propriétés objectives, au sens où semblait l’entendre Russell : pas de propriétés totalement indépendantes de nous, sur au moins deux niveaux.

A un premier niveau, la physique est la mise en relation : la position de telles aiguilles, une certaine itération du déplacement d’une règle, etc, avec la manipulation de symboles sur une feuille ou un tableau, et avec une histoire. Cette mise en relation est le fonctionnement d’un certain système physique. Pour le physicien il n’y a pas de “propriété objective devant nous”, comme cela est évident en particulier en physique quantique : le dispositif expérimental est constitutif du système physique de mise en relation, il est constitutif de la propriété. La notion de propriété indépendante de tout

dispositif est un leurre, une confusion avec le fait que le fonctionnement de différents dispositifs peuvent se trouver en correspondance systématiquement cohérente (l'objectivité interpersonnelle), et avec le fait que la potentialité d'un certain fonctionnement est suffisante pour parler d'une propriété de l'environnement (l'objectivité apersonnelle).

A un second niveau, il est absurde de croire qu'il existe une conception canonique du monde : nous ne pouvons parler que du monde tel que nous le concevons, avec le système de catégories et de dérivation logique qui est le notre. Il n'y a pas d'autre discours signifiant pour nous qu'un discours qui dépend de nous par ces choix de catégories et de structure logique ([Putnam, 1987](#)) :

The situation does not itself legislate how words like "object", "entity", and "exist" must be used. What is wrong with the notion of objects existing "independently" of conceptual schemes is that there are no standards for the use of even the logical notions apart from conceptual choices.

L'idée est que la remise en cause d'un réalisme simpliste doit pousser à repenser d'une manière plus adéquate des concepts tels que la connaissance, la perception et l'expérience sensorielle. Pour moi, la valeur de l'approche sensorimotrice est de faire une tentative dans cette direction.

Deuxième partie

L'espace de l'expérience

Chapitre 5

Espace réel, espace sensorimoteur

Lorsque nous nous écartons de l'échelle humaine, les propriétés spatiales du monde se trouvent aller à l'encontre de nos intuitions les plus profondes. Les théories développées pour rendre compte des phénomènes typiques des très grandes ou très petites échelles se trouvent ainsi contraintes d'inscrire ces phénomènes dans des espaces ayant un nombre élevé de dimensions, des discontinuités systématiques ou accidentelles, des symétries inattendues, etc.

Prétendre que notre expérience de la spatialité est une forme de savoir pratique sur la réalité dans laquelle nous vivons, une réalité qui aurait la structure d'un espace tridimensionnel, régulier, euclidien, peuplé d'objets possédants certaines propriétés de localisation et d'opacité, ne coule donc pas de source. La physique du monde, en effet, peut être retenue contre ce réalisme naïf : le monde est multiforme selon les échelles et les phénomènes, et la "réalité" précédemment décrite n'est pas plus un absolu dans l'ensemble du monde physique que la terre n'est au centre de l'univers. Le réalisme doit s'appuyer au minimum sur la considération d'une échelle et d'un champ des phénomènes déterminés par notre fonctionnement en tant que structure physiologique : paradoxalement, le réalisme ne peut pas être indépendant de nous.

Cela, toutefois, ne signifie pas que la réalité précédemment décrite soit un concept vain et subjectif. L'espace tridimensionnel, approximativement euclidien, se trouve être en particulier l'espace des artisans géomètres : une structure objective de manipulations

et déplacements que peut opérer l’homme en tant que système sensorimoteur. Sauf à la considérer comme un remarquable hasard, cette coïncidence de nos rapports verbaux sur l’espace avec les caractéristiques d’une structure de capacités pratiques suggère que l’espace de l’expérience n’est pas indépendant de nos capacités d’interaction sensorimotrice : pas plus “à l’extérieur” de nous, dans une réalité conçue hors de toute référence à l’humain, que “en nous”, dans une contrainte représentationnelle purement interne.

Dès lors, prétendre que “la spatialité de l’expérience visuelle est dans la structure de nos interactions visuo-motrices” ne peut être tenu pour particulièrement extravagant. Mais que pourrait signifier, précisément, la notion de spatialité d’un ensemble d’interactions sensorimotrices ?

5.1 Les fondements de la géométrie

A l’occasion d’un débat différent, celui concernant les fondements empiriques de la géométrie, Helmholtz puis Poincaré ont indirectement formulé il y a plus d’un siècle leur opinion sur une fondation sensorimotrice de l’espace¹.

Helmholtz cherchait à baser la géométrie sur la manipulation d’un ensemble d’objets particuliers, les objets rigides, proposant ainsi une vision très opérationnelle de l’espace (von Helmholtz, 1876). Sa théorie reposait de manière critique sur la possibilité de déplacer librement ces objets, et de les faire tourner autour d’un de leurs points. Cette position le conduira à la notion d’espace de courbure constante, un cas particulier de la théorie développée au même moment par Riemann et dont il ne prend connaissance qu’après avoir poussé assez loin ses propres réflexions (Boi, 1995). Pour Helmholtz, l’espace physique est la structure de la manipulation des objets rigides.

Poincaré, s’il faisait également jouer un rôle critique aux objets rigides, le faisait d’une manière plus détournée. Il proposait une genèse de l’espace basée sur la possibilité de compenser, vis-à-vis de nos impressions visuelles, les mouvements d’un objet par un mouvement inverse de notre tête (Poincaré, 1895). Il décrivait comment un indi-

¹La fin du XIXe siècle était le théâtre d’un questionnement en profondeur sur l’espace. La maturation du concept de géométrie non-euclidienne à partir du milieu du siècle avait renversé le statut métaphysique donné par la philosophie kantienne à l’espace euclidien : l’homme démontrait sa capacité à concevoir d’autres espaces. Il se posait, en conséquence, la question de l’origine du statut particulier de la géométrie euclidienne dans l’histoire humaine, une géométrie sur les devants de la scène pendant plus de deux millénaires. Il se posait également la question du statut approximatif ou rigoureux de la géométrie, dès lors qu’elle n’était plus l’a priori synthétique kantien. Et ces questions se posaient avec une insistance exceptionnelle : elles remettaient en cause tout un pan de notre conception de l’entendement.

vidu, confronté à des impressions sensorielles et des sensations musculaires qui n'auraient au départ aucune signification particulières pour lui, devrait pouvoir construire successivement les notions de déplacements, de leur composition, de dimension de l'espace, de point, etc. Pour Poincaré, l'espace est une symétrie dans la structure des dépendances qui existent entre nos impressions sensorielles, nos impressions musculaires, et la configuration d'un environnement composé d'objets rigides.

La notion de groupe de transformations étant déjà apparue comme le déterminant fondamental d'une géométrie, le programme d'Erlangen date de 1872, les deux propositions s'articulent autour de cette notion, et consistent en fait à trouver une activité sensorimotrice dont la combinatoire fait d'une manière ou d'une autre surgir un groupe.

Helmholtz et Poincaré étaient loin d'être les seuls à préférer à l'a priori kantien une vision essentiellement empirique de l'espace : à la même époque, de nombreux mathématiciens et physiciens, tels Clifford ou Hertz, défendaient l'idée d'une géométrie décrivant la structure du monde plutôt qu'une "intuition de l'esprit". Les discours de Helmholtz et de Poincaré, cependant, avaient comme spécificité leur référence très claire à des capacités sensorimotrices plutôt qu'à quelques phénomènes seulement accessibles aux physiciens. Poincaré, en particulier, en affirmant (Poincaré, 1902) :

Aucune de nos sensations, isolée, n'aurait pu nous conduire à l'idée de l'espace, nous y sommes amenés seulement en étudiant les lois suivant lesquelles ces sensations se succèdent.

ançrait l'espace dans notre corps plutôt que dans les propriétés d'un monde conçu hors de référence à l'humain, et il l'ançrait même spécifiquement dans les dépendances sensorimotrices.

La position sensorimotrice de Poincaré et de Helmholtz, cependant, est dans une certaine mesure ambiguë : leurs textes laissent l'impression que pour eux il ne doit y avoir qu'un espace, et que, tout en étant sensorimoteurs, leur question était bien de savoir quelle était "vraiment" la géométrie du monde.

La question de la structure des interactions sensorimotrices, en comparaison, est plus pragmatiquement de comprendre ce qui pourrait justifier l'utilisation du qualificatif "spatial" pour caractériser certaines dépendances sensorimotrices, et de justifier certaines limites dans les possibilités d'analogies entre différents ensembles de dépendances. La question *n'est pas* celle de la spatialité d'un monde conçu sans référence à l'activité potentiellement déployable par un système.

Il reste que les interrogations de Poincaré et de Helmholtz sont une source d’inspiration directe pour mener une réflexion sur la structure spatiale des interactions sensorimotrices, puisque montrer comment une géométrie peut être construite à partir de l’analyse de certaines interactions est un indicateur du type d’interactions cohérentes avec l’utilisation du terme “spatial”.

5.2 La tradition sensorimotrice de l’espace

Il y a en fait une longue tradition de théories sensorimotrices sur la sensation de spatialité de l’expérience visuelle, développées par les philosophes, psychologues et physiologistes dits “empiristes”. Locke, Condillac, Berkeley, Mill, Bain, Lotze argumentaient déjà qu’il n’y a dans la sensation de spatialité rien d’autre que des jugements, qui découlent de l’expérience répétée de certaines dépendances entre les informations sensorielles et les actions motrices : position de l’œil, accommodation et convergence, manipulation des objets.

John Stuart Mill, par exemple, soutenait que la notion d’étendue est profondément liée aux mouvements de l’œil, et que c’est ce qui nous trompe à concevoir l’étendue comme un phénomène de synchronisme, plutôt que de succession (Mill, 1865) – une idée qui a retrouvé un nouveau souffle au regard de la cécité au changement.

Le philosophe et psychologue Alexander Bain, quant à lui, suggérait que, lorsque nous disons de deux points A et B qu’ils sont séparés par un espace, nous voulons simplement dire qu’il y a une série de sensations musculaires entre le moment où notre doigt part de A et le moment où il arrive à B. Selon lui, c’est la durée plus ou moins longue d’effort musculaire qui nous donnerait la sensation d’étendue (Bain, 1868).

A peu près à la même époque, Rudolf Hermann Lotze développait dans une perspective différente sa théorie des “signes locaux”, l’idée que chaque fibre du nerf optique est associée avec un signe qui identifie la localisation spatiale de la stimulation, un signe qu’il considère établi par l’expérience sensorimotrice (Lotze, 1894). Helmholtz reprendra d’ailleurs cette idée, en argumentant que le signe local doit résulter de la corrélation existant, lors du mouvement des yeux, entre les signaux délivrés par les fibres du nerf optique, du fait du recouvrement spatial des champs réceptifs des capteurs de la rétine (von Helmholtz, 1867). L’idée sera reprise une nouvelle fois, dans un cadre plus contemporain, par Koenderink (1984), puis récemment dans divers tra-

vaux de robotique développementaliste (Pierce and Kuipers, 1997; Olsson et al., 2006; Kaplan and Hafner, 2005).

* * *

Piaget, inspiré par Poincaré, poursuivra les idées des empiristes à partir d'un nouveau point de vue. Il remplace le concept d'association entre stimulations sensorielles et commandes motrices par celui de *schèmes* sensori-moteurs, en insistant sur la structure de combinaison de ces schèmes (Piaget, 1937) :

Le premier des schèmes constitutifs de l'espace enfantin est celui que Stern a désigné du nom d'espace buccal. Les déplacements de la bouche par rapport aux objectifs à sucer ou des objectifs par rapport à la bouche constituent donc les "groupes" pratiques les plus simples qu'il soit possible d'observer chez l'enfant. [...] nous n'hésiterons pas à parler, avec Poincaré, de "groupes" *pour désigner les conduites mêmes de l'enfant*, dans la mesure où elles sont capables de réversibilité, ou de corrections les ramenant au point initial.

Il est important de noter que le discours de Piaget est en réalité en décalage notable par rapport à la position de Poincaré. Ce dernier décrit en effet une conscience ayant des "impressions" sensorielles et musculaires, et capable d'acquérir un savoir sur certaines caractéristiques de la relation existant entre ces impressions. L'espace de Poincaré, intéressé par la géométrie, est un *savoir sur une certaine structure*. Piaget présente, lui, le tableau d'une structuration des interactions sensorimotrices en comportements de plus en plus complexes. L'espace de Piaget est un *savoir faire ayant une certaine structure* :

Au total, l'espace buccal est un espace pratique, qui permet à l'enfant de retrouver des positions, d'exécuter des mouvements, de s'adapter à des formes et à des dimensions [...]

Il n'y a pas d'impressions sur lesquelles sont construites un savoir, il y a la construction d'impressions au travers du développement d'un savoir faire, d'un savoir s'engager dans différentes dynamiques sensorimotrices.

Dans l'approche de Piaget, ce n'est qu'à la toute fin du développement qu'une forme de savoir sur, c'est-à-dire de capacités déductives et démonstratives (au sens

de “pointer”), est établie par la capacité de coordonner différents schèmes et de les exploiter dans des situations inconnues :

La nouveauté du [troisième] stade, c’est la coordination des différents groupes pratiques entre eux, donc de l’espace buccal avec l’espace visuel, de l’espace visuel avec l’espace tactile kinesthésique, etc. [...] Le type de conduites qui est au point de départ des manifestations du quatrième stade est l’application des moyens connus aux situations nouvelles.

Avant le troisième stade, les caractéristiques spatiales des mouvements sont là, mais sans être identifiées en tant que telles, l’idée étant que cette identification est la capacité de coordonner différents schèmes sensorimoteurs et d’entretenir des croyances quant au déroulement de ces schèmes dans une variété de situations².

Piaget est également célèbre pour son argumentation d’un rôle critique de la locomotion dans l’émergence de nos compétences spatiales.

Par exemple, selon sa théorie, l’enfant part d’un mode de fonctionnement égocentrique – c’est-à-dire que, ayant trouvé un objet sur sa droite, il va continuer à le chercher sur sa droite après avoir bougé d’une manière qui rende cette attitude inappropriée –, et ne développe qu’ensuite, grâce à la locomotion, une capacité d’appréhender l’espace de manière allocentrique, c’est-à-dire tenant compte de la position des objets et de lui-même dans un cadre de référence extérieur (Piaget and Inhelder, 1972).

Ou bien, reprenant l’idée largement favorisée par les empiristes d’une éducation de la vue par le toucher, Piaget argumente que la perception de la profondeur requiert la coordination de la vision avec la faculté d’atteindre, et que pour le jeune enfant il existe donc un “espace proximal” délimité par l’accessibilité (Piaget, 1937). Au delà de cette limite, selon l’image de Piaget, l’enfant pré-locomoteur perçoit le monde comme nous percevons les étoiles sur la voute céleste, et ce n’est que grâce à la locomotion que l’enfant apprend à désambiguer “l’espace distant”, à estimer les distances et la taille des objets. Un tel apprentissage serait basé sur la calibration de l’information visuelle en termes d’unités motrices, la distance d’un jouet étant le nombre de mouvements locomoteurs répétés nécessaire pour arriver à être en position de s’en saisir (Kaufman, 1974).

²On pourrait penser que les positions de Poincaré et de Piaget peuvent être réconciliées par la formule que le savoir faire de l’enfant demande une forme de savoir sur du cerveau. Cette idée n’est sans doute pas absente de l’approche présentée dans Philipona et al. (2003), ou de l’histoire de l’équipe d’ingénieurs évoquées dans O’Regan and Noë (2001). Avec le recul, comme je le détaille plus en section 6.4, il me semble toutefois que cette métaphore attirante du cerveau qui “sait” n’apporte que des confusions.

Bien que toujours enseignée, l'approche de Piaget n'a plus le crédit qu'elle avait il y a quelques décennies. Les travaux des successeurs de Piaget, en effet, n'ont jamais permis de mettre à jour la théorie formelle du développement que ce dernier a cherché pendant sa vie (Bates et al., 1998), mais surtout ses idées, comme plus généralement celles des empiristes, sont tenues pour avoir été contredites par les nombreuses capacités précoces des nouveaux-nés (Newcombe, 2002).

S'il est établi en effet que les débuts de la locomotion volontaire induisent en règle générale une transition critique au niveau des capacités de l'enfant (Campos et al., 2000), – capacités perceptives, de cognition spatiale, de communication distale –, de nombreuses capacités sont observables avant même qu'une quelconque calibration de l'information sensorielle par les capacités motrices puisse avoir pris place.

A titre d'exemple, les nouveaux-nés montrent certaines capacité d'identification de la taille des objets malgré leur distance et leur inclinaison (Slater et al., 1990), alors que leurs capacités de faire des mouvements en direction de ces objets, de contrôler l'accommodation et la convergence de leur système oculaire, de se mouvoir, sont encore extrêmement limitées. D'autre part, les enfants semblent comprendre très jeunes que les personnes observant une scène d'un autre point de vue voient quelque chose de différent (Newcombe, 2002), ce qui contredit l'image d'enfants fonctionnant à l'origine sur un mode exclusivement égocentrique.

Enfin, le rôle de l'expérience locomotrice n'est pas clair : permet-elle le développement d'un codage relatif de la position des objets ou la mise à jour d'un codage égocentrique par un suivi des déplacements, habitue-t-elle les enfants à la disparition des objets du champ de vision (du fait des occlusions) en leur apprenant garder un oeil à l'endroit de la disparition, ou est-elle surtout importante parce qu'elle nécessite de garder un but à l'esprit ?

* * *

Les conceptions sensorimotrices de l'espace ont été mises en avant par les empiristes, dont l'objectif était de défendre l'idée que l'espace peut être *appris*. Dans ce contexte, il est important de préciser que l'approche sensorimotrice présentée dans la suite n'est pas une approche développementale : il ne s'agit pas de proposer une évolution ou un enchaînement de compétences, mais de chercher à analyser certaines caractéristiques essentielles des interactions sensorimotrices dans lesquelles nous pouvons

nous trouver engagés, afin d'expliquer certaines limites fondamentales sur la possibilité de constater certaines analogies entre différentes situations sensorimotrices.

C'est une chose de dire que l'espace est dans la structure des interactions sensorimotrices, c'en est une autre d'affirmer que les compétences sensorimotrices peuvent ou doivent émerger d'une *tabula rasa*, quoique cela puisse vouloir dire. Pour faire une analogie : on peut distinguer les questions de la structure du raisonnement humain (la structure des interactions sensorimotrices), de l'apprentissage d'une manière de raisonner (de l'apprentissage des compétences sensorimotrices et des analogies de situations), et du point de départ de cet apprentissage.

5.3 Différents types de dépendances sensorimotrices

Les interactions sensorimotrices dans lesquelles un organisme peut se trouver engagé résultent, d'une part, de la relation induite par son environnement entre la stimulation sensorielle et les commandes motrices, et, d'autre part, de ses réflexes moteurs et de son contrôle rationnel de l'action (Fig. 5.1). Je parlerai dans la suite de *dépendance sensorimotrice* pour référer à la première partie des interactions, et de *contrôle sensorimoteur* pour référer à la seconde.

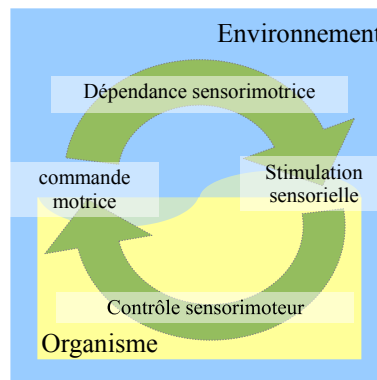


FIG. 5.1 – Interactions sensorimotrices

Un organisme peut se trouver confronté à une variété de dépendances entre stimulations sensorielles et commandes motrices, induites par différents environnements : différentes scènes visuelles, différents environnements sonores, etc. Un organisme peut mettre en oeuvre une variété de contrôles sensorimoteurs, en fonction de différents objectifs, de la mémorisation de différents faits, etc. Les interactions sensorimotrices

dans lesquelles un organisme peut se trouver engagé sont l'ensemble des évolutions résultants du couplage de ces variétés de dépendances et de contrôles.

On dira que deux systèmes peuvent s'engager dans une même structure d'interactions sensorimotrices s'il est possible d'établir une relation biunivoque des stimulations et commandes motrices de l'un sur les stimulations et commandes motrices de l'autre, et que cette relation fait correspondre l'ensemble des évolutions possibles de l'état sensorimoteur des deux systèmes. Il s'agit de la formalisation de l'idée sous-jacente à la substitution sensorielle : celle que le TVSS (Sec. 3.2) induit une dépendance entre mouvements de la tête et stimulations tactiles qui vérifie les mêmes propriétés que la dépendance entre mouvements de la tête et stimulations rétinienne, c'est-à-dire que, modulo une mise en correspondance fixe entre certaines stimulations tactiles et certaines stimulations rétinienne, les dépendances sont les mêmes pour l'ensemble des environnements, et que si le sujet parvient à développer au travers du TVSS des capacités de contrôle sensorimoteur similaires à celles mises en oeuvre dans la vision, alors il peut s'engager dans une structure d'interactions sensorimotrices identique à celle typique de l'appareil visuel.

En toute rigueur, il n'y a pas de sens à parler de structure des interactions sensorimotrices sans tenir compte à la fois des parties dépendance et contrôle. Un même système de dépendances couplé à différents répertoires de contrôle, tel par exemple que dans le cas d'un jongleur débutant et d'un jongleur expert, conduit à des structures d'interactions sensorimotrices différentes. Toutefois, les dépendances peuvent imposer quelques contraintes fondamentales sur les interactions, indépendamment du contrôle : ainsi, faire un tour sur soi-même, quelque soit le détail du contrôle moteur mis en oeuvre pour cela, nous retrouve dans la situation perceptive initiale, ou fermer les paupières, quelque soit le contrôle mis en oeuvre, nous laisse dans une situation où les commandes motrices ne sont plus accompagnées d'un changement de stimulation rétinienne. C'est de ce type de contraintes élémentaires que parle la suite de ce chapitre.

* * *

Selon les systèmes sensorimoteurs considérés, il est naturel de s'attendre à divers types de dépendances sensorimotrices. Les deux types fondamentaux suivants, caractérisées par la forme de la dépendance dans un environnement fixe, paraissent en particulier d'intérêt :

1. chaque commande motrice est associée à une stimulation sensorielle déterminée (e.g. la dépendance, dans un environnement fixe, entre la stimulation rétinienne et l'innervation des muscles oculaires) : il y a une relation *associative* de la forme $s = \varphi(m)$ entre les commandes motrices m et les stimulations s ,
2. chaque variation de commande motrice est associée à une variation de la stimulation sensorielle (e.g. la dépendance, dans un environnement fixe, entre la stimulation rétinienne et les commandes locomotrices) : il y a une relation *différentielle*, par exemple de la forme $\dot{s} = \varphi(s, m, \dot{m})$, entre les commandes motrices m et les stimulations s .

Ces deux types de dépendance renvoient à des situations très différentes : par exemple, dans le premier cas, une trajectoire motrice $t \in [0, T] \mapsto m(t)$ résulte en une stimulation sensorielle $s(T)$ qui ne dépend pas de l'historique de la trajectoire, mais seulement de la dernière commande motrice $m(T)$, ce qui, en règle générale, n'est pas vrai dans le second. Pour l'un ou l'autre de ces types de dépendance, il se pose donc la question des caractéristiques pouvant mériter le qualificatif de spatial.

La section 6 analyse le cas des dépendances de type associatif. Elle fonde la notion de spatialité sur l'existence de certaines modifications de l'environnement, tel que le mouvement d'un objet sur un fond uniforme, qui sont associées à des changements de dépendance compensables par une adaptation adéquate des commandes motrices. Comme il a déjà été dit, c'est à partir de ce genre d'argumentation que Poincaré a construit sa genèse de la géométrie euclidienne, dans un cadre conceptuel toutefois différent, et sans véritablement pousser, à ma connaissance, de formalisation mathématique de cette idée (c.f. Annexe 7).

La notion de spatialité des dépendances différentielles est en comparaison plus classique et beaucoup mieux établie. Elle peut être rappelée en quelques lignes ci-dessous. Pour guider l'intuition, on peut noter que la robotique donne des exemples de dépendances différentielles ayant une structure mathématique particulièrement régulière, par exemple le cas d'un véhicule à deux roues doté d'une caméra, se déplaçant sans glissement sur un sol planaire dans un environnement fixe (Fig. 5.2).

Si l'on note $\omega \in \Omega$ la configuration du système global, la notion d'espace des déplacements du véhicule repose alors, pour le mathématicien, sur trois caractéristiques interdépendantes³ :

³Le concept d'espace évoqué ici repose implicitement sur l'idée d'un fibré principal dont la base est la

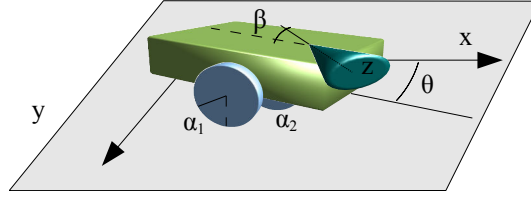


FIG. 5.2 – Véhicule mobile. La description du système global implique une configuration relative (l’angle des roues, de la caméra, le zoom de la caméra), et des variables de position du véhicule (la position et l’orientation dans un repère).

- *un concept de corps déplaçable*, c’est-à-dire de corps qui peut être dans des configurations relatives identiques mais en des endroits différents. Dans le cas du véhicule, la configuration globale du système est caractérisée par des variables indiquant l’angle des roues, de la caméra, le zoom de la caméra $\omega_i = (\alpha_1, \alpha_2, \beta, z) \in \Omega_i$, et par des variables de position $\omega_e = (x, y, \theta) \in \Omega_e$, et cela fait partie de la conception du problème que l’ensemble des possibles est de la forme $\Omega = \Omega_e \times \Omega_i$.
- *une loi de déplacement du corps* : l’existence d’équations reliant changements de configuration relative et changement de configuration globale. Dans l’exemple du véhicule, il s’agit d’équations de la forme $\dot{\omega}_e = A(\omega_e, \omega_i)\dot{\omega}_i$ (c.f. par exemple [Kelly and Murray \(1994\)](#); [Marsden and Ostrowski \(1998\)](#)).
- *l’invariance de cette loi en différentes localisations* : le fait que la loi de déplacement est la même quelque soit la position du corps déplaçable, c’est-à-dire qu’il est possible de trouver une manière de mettre en correspondance les voisinages de chaque couple de positions (ω_e, ω'_e) par un difféomorphisme $T : \mathcal{V}(\omega_e) \rightarrow \mathcal{V}(\omega'_e)$ vérifiant $A(T(\xi), \omega_i) = DT(\xi)A(\xi, \omega_i)$ pour tout $(\xi, \omega_i) \in \mathcal{V}(\omega_e) \times \Omega_i$ (les équations de déplacement aux voisinages de différentes positions sont les mêmes, modulo la bonne mise en correspondance des positions indiquées dans ces équations).

Dans le cas, certes irréaliste mais tout de même intéressant comme élément de référence, où l’état global du système est indiqué de manière régulière et non ambiguë par les variables sensorielles extéroceptives et proprioceptives, et où la configuration

posture du corps et la fibre sa localisation spatiale, dans l’esprit par exemple de [Radford and Burdick \(1998\)](#), plutôt que sur la notion d’espace Riemannien. Cela a pour conséquence de “restreindre” l’approche aux espaces symétriques. On peut toutefois considérer qu’il s’agit d’un modèle local, et qu’il doit se poser à partir de cela la question du rapièçage de différents modèles locaux – tout comme le fait l’approche Riemannienne à partir d’un modèle local d’espace euclidien (donc symétrique).

relative est déterminée par les variables motrices, on voit à partir de ce qui précède que l'espace est l'existence de certaines correspondances entre les changements de configurations sensorielles et motrices.

Les travaux présentés en section 8 se placent dans le cadre général de ce style de dépendances différentielles, afin de proposer une notion de similarité sensorielle qui découle des dépendances dans lesquelles un dispositif d'acquisition des données peut s'engager, plutôt que de découler de la similarité des *représentations* des données sensorielles. La notion de similarité résultante est exploitée dans le contexte d'une technique d'analyse non supervisée des données, la réduction de dimension ([Edelman and Intrator, 1997](#); [Tenenbaum, 1998](#); [Tenenbaum et al., 2000](#); [Roweis and Saul, 2000](#)).

Chapitre 6

Dépendances sensorimotrices associatives

Quelle signification pourrait avoir la notion de spatialité d'une dépendance sensorimotrice associative ? Les propriétés d'une dépendance de la forme $s = \varphi(m)$ entre des variables sensorielles et motrices multidimensionnelles peuvent-elles réellement caractériser la dimension de l'espace dans lequel l'organisme évolue ? Peuvent-elles caractériser les différents types de déplacements dans cet espace ? En premier lieu, peuvent-elles simplement distinguer entre ces changements particuliers que sont les déplacements d'un objet de l'environnement, et ces autres changements que sont les changements d'intensité sonore ou lumineuse ?

Qu'il puisse y avoir une réponse positive à ces questions n'est en aucune manière évident¹. On peut naturellement douter de l'idée que les propriétés intrinsèques d'une dépendance de la forme $s = \varphi(m)$, i.e. les propriétés indépendantes de la représentation des variables sensorielles et motrices, permettent de caractériser la dimension d'un espace sous-jacent ou de fonder une distinction très élaborée entre différents types de modifications de la dépendance. En effet, il est aisé de se convaincre que les seules propriétés intrinsèques d'une dépendance associative sont, d'une part, le nombre de variables motrices, et, d'autre part, les propriétés d'équivocité de la dépendance – les

¹Ce ne l'était en tout cas pas pour moi, c'est d'ailleurs la raison pour laquelle mon état d'esprit, au cours des travaux décrits dans ce qui suit, était que tout argument suggérant qu'il existe une réponse est bon à prendre. Cela explique une certaine facilité à considérer, en première approche, des hypothèses d'un réalisme assez approximatif.

cas où différentes valeurs des variables motrices sont associées à une même valeur des variables sensorielles.

Cela apparaît clairement lorsqu'on se restreint au cas particulier de dépendances sensorimotrices qui sont linéaires dans une certaine représentation des variables sensorielles et motrices, i.e. de la forme $s = Am$ dans la bonne représentation. Dans un tel cas, en effet, on voit de manière transparente que les seules propriétés intrinsèques de ces dépendances sont la dimension de l'espace (vectoriel) moteur et la dimension du noyau², des propriétés pour le moins rudimentaires et effectivement incapables d'apporter une réponse aux questions évoquées.

Par contre, un changement significatif intervient lorsqu'on considère qu'il ne s'agit pas d'analyser "dans l'absolu" les propriétés d'une dépendance $s = \varphi(m)$, hors de tout contexte, mais de *la mettre en correspondance avec une variété définie d'autres dépendances*, résultant d'un ensemble de configurations possibles de l'environnement. De ce point de vue, on peut espérer fonder une distinction catégorique entre différents changements de dépendances, selon la manière dont les dépendances associées peuvent être mises en correspondances les unes avec les autres au travers d'un répertoire d'adaptations sensorimotrices. Les deux articles suivants présentent une tentative d'avancer dans cette direction.

Quelques compléments sont proposés ensuite – venant d'abord plutôt soutenir l'approche, en section 6.3, puis plus critiques, en section 6.4. Enfin, la section 6.5 présente les grandes lignes d'une reformulation de l'approche qui, tout en gardant des limitations que j'essaie de détailler, me paraît plus cohérente, et d'une certaine manière plus sensorimotrice.

En préliminaire des articles, il faut noter que l'analyse pointe l'*existence* de certaines possibilités non triviales au niveau des interactions sensorimotrices, du fait des propriétés des dépendances sensorimotrices, mais ne propose pas de contrôles pour réaliser ces possibilités. Ainsi, le contrôle sensorimoteur effectivement mis en oeuvre pour poursuivre une cible, i.e. pour compenser son mouvement, est a priori basé sur un feedback sensoriel, alors que l'analyse proposée ne s'intéresse qu'à l'existence d'une commande motrice permettant cette compensation. Comme on l'a déjà dit, il

²C'est-à-dire : ces propriétés peuvent être déterminés indépendamment de la représentation des variables, et si ces propriétés sont égales pour deux dépendances, alors cela signifie que l'on considère une même dépendance dans des représentations différentes des variables sensorielles et motrices (on peut trouver un changement de représentation des variables telles que les deux dépendances soient identiques).

s'agit uniquement d'analyser certaines *contraintes* existant sur les interactions sensori-motrices.

6.1 Article : Is There Something Out There ?...

L'article "*Is There Something Out There ? Inferring Space from Sensorimotor Dependencies*" est focalisé sur la notion de dimension. Il considère le cas d'un organisme contrôlant un grand nombre de capteurs articulés dans un environnement extrêmement simplifié, dont chaque configuration impose une dépendance associative entre les stimulations sensorielles et les commandes motrices. L'article montre qu'il est possible de distinguer certains mouvements du dispositif ou de l'environnement associés aux modifications du système {organisme, environnement} laissant les stimulations sensorielles invariantes. Dans l'exemple considéré, ces mouvements correspondent aux mouvements rigides (du système sensoriel ou, de manière symétrique, de l'environnement), et la dimension du groupe de symétries de l'espace sous-jacent aux dépendances apparaît donc comme étant effectivement une propriété de leur ensemble.

Is There Something Out There? Inferring Space from Sensorimotor Dependencies

D. Philipona

david.philipona@polytechnique.org
Sony CSL, 75005 Paris, France

J.K. O'Regan

oregan@ext.jussieu.fr
Laboratoire de Psychologie Expérimentale, CNRS, Université René Descartes,
92774 Boulogne-Billancourt Cedex, France

J.-P. Nadal

Jean-Pierre.Nadal@lps.ens.fr
Laboratoire de Physique Statistique, Ecole Normale Supérieure,
75231 Paris Cedex 05, France

This letter suggests that in biological organisms, the perceived structure of reality, in particular the notions of body, environment, space, object, and attribute, could be a consequence of an effort on the part of brains to account for the dependency between their inputs and their outputs in terms of a small number of parameters. To validate this idea, a procedure is demonstrated whereby the brain of a (simulated) organism with arbitrary input and output connectivity can deduce the dimensionality of the rigid group of the space underlying its input-output relationship, that is, the dimension of what the organism will call physical space.

1 Introduction ---

The brain sits inside the cranial cavity monitoring the neural signals that come into it and go out of it. From this processing emerge the notions of self, outside space, objects within that space, and object attributes like color, luminosity, and temperature. Even simple organisms that have little or no cognitive ability clearly possess these concepts at least implicitly, since they show spatially adapted behavior like locomotion, navigation, grasping, and discrimination of different objects.

How is this possible? What kind of algorithms must be at work inside biological brains for these notions to be extracted from the neural activity in a mass of unlabeled nerve fibers? Do brains have this capacity because phylogeny has yielded a brain structure that is specially adapted to understanding the notion of space?

Here we investigate a more radical hypothesis: the possibility that what brains do is to continuously calculate statistics on their neural activity in an attempt to characterize it using a small number of parameters. We take the extreme case where the brain has absolutely no a priori information about outside physical space (whether it exists at all, whether it has a metric, whether it is Euclidean, how many dimensions it possesses). We assume that the linkage between motor commands and the resultant motions of the organism is unknown to the brain and totally arbitrary. We further assume that the brain has no information about what nerve pathways correspond to sensors sensing internal or external states.

We show that there is a simple procedure that a brain can use to arrive at a distinction between the body, which it can control, and the outside world, which it cannot completely control. There is a simple algorithm that leads to a characterization of the number of variables necessary to describe the organism's body (insofar as the body states affect incoming sensory information).

Further, and most interesting, we show that the brain can deduce the dimensionality of outside physical space and the number of additional non-spatial parameters needed to describe the attributes of objects or entities within it.

Our hypothesis is therefore that what biological organisms perceive as being the limits of their bodies, as well as the geometry and dimensionality of space outside them, are deducible, without any a priori knowledge, from the laws linking the brain's inputs and outputs. The approach we are taking derives from the basic idea that the basis of sensory experience consists in extracting and exercising laws of sensorimotor dependencies (O'Regan & Noë, 2001).

2 A Simple Organism

Let us imagine a simple organism consisting of an articulated arm fixed to the ground (see Figure 1). At the end of each of its fingers is an eye composed of a number of light-sensitive sensors. Imagine in addition that the organism has proprioceptive devices that signal the position of the different parts of the arm. The environment, we shall suppose, consists of a set of lights. The signals provided by the sensors are transmitted to a brain that controls effectors that move the arm.

Let us suppose that the brain has no a priori knowledge about the body that it is connected to and that this body is the sole connection it has to the outside world. What comprehension of this "exterior" can the brain attain, and how can it be obtained?

Certainly the simplest thing the brain can do is to start by trying various motor commands at random. Gradually, the brain will notice that it can make a distinction among the mass of sensory inputs that it receives. It will notice that certain inputs, or certain combinations of inputs, always react

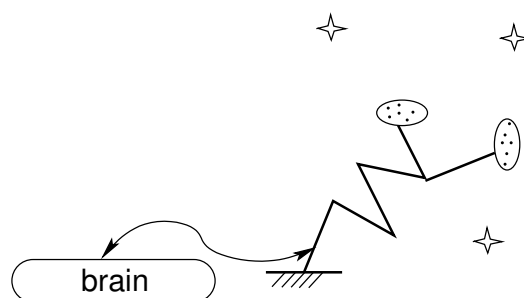


Figure 1: A simple organism consisting of an articulated arm with two “fingers” and a composite “eye” mounted on each.

the same way to motor commands, while the other inputs show only partial, unsystematic relations to motor commands. What is the natural conclusion that the organism can deduce from this fact? It is the fact that its universe can be separated into a part that the organism can completely control and a part that the organism can only partially control.

We shall call the first part, over which it has complete control, the organism’s *body* and the second part the organism’s *environment*. We shall call the first type of inputs *proprioceptive* and the others *exteroceptive*.¹ We shall say the *body is stationary* when proprioception is constant, and we shall say the *environment is stationary* when exteroception is constant.² Note that since the organism is totally naïve about its environment (and even about the fact that there is such a thing at all as an environment), it has no choice but to define these notions.

The brain can now attempt to understand its environment. Changes that occur in exteroceptive sensors when the body is stationary can be taken to derive from the environment. We shall assume that the brain attempts to account for these changes, which are defined in the very high-dimensional space of the number of exteroceptors, in terms of a much smaller number of parameters. For example, in the case of the articulated arm, there may be 40 photoreceptors, but their outputs are completely determined by a much smaller number of parameters: the positions of the three lights in

¹ We follow the terminology used in Kandel, Schwartz, and Jessell (2000) and stress that exteroceptive sensors are sensitive not only to changes of the environment but to motion of the body as well, while proprioceptive are sensitive to changes of the body only (which is more restrictive than the usual use of this term). Also, it should be noted that this distinction arises gradually: certain inputs that might at first seem completely determined by motor commands will later turn out in fact only to be partially determined by them.

² This definition is compatible with sensors sensitive to derivatives of position, since, for example, if both velocity and position are constant, then velocity must be zero.

the environment. We shall call the vector of this reduced parameter set a *representation of the state of the environment*.

The brain can also issue motor commands. When it does this sufficiently quickly, it can assume that most of the resulting changes in its exteroceptive input will be due to its own motion and not to motion of the environment. Again, a low-dimensional parameter set can be extracted from the exteroceptive changes that occur in this case, and we shall call the resulting vector a *representation of the exteroceptive body*.

The representation of the exteroceptive body is richer than the representation of the proprioceptive body in the sense that it provides a way for the organism to discover the relationship between its body and the state of the environment. Importantly, the organism can note that certain exteroceptive changes caused by motion in those body parts that bear sensors can be compensated by particular environmental changes. For example, in the example of the articulated arm, when the set of photoreceptors is displaced rigidly by an arm motion, there is a corresponding inverse rigid displacement of the environmental lights that can be made, which will cause exteroception to return to its original value. Note, however, that such compensations are possibilities that need never actually occur for our algorithm to work. It is unlikely that the environment would ever actually move in this particular rigid fashion. Note also that the notion of compensability derives in an intrinsic fashion from the fact that the response of a single set of sensors, namely the exteroceptive ones, varies as a function of two distinct sources of variation: body changes and environmental changes. Proprioceptive sensors, since they are sensitive only to a single source of variation body changes, do not provide information about this kind of compensability.

The notion of compensability, arising from the confrontation of two distinct sources of change within a single set of inputs, has fundamental implications for the genesis of the notion of space in an organism:

- Compensability defines a class of body movements with a particular structure. If the body makes such a movement, then the movement bringing the body back to its original position is also compensable. If the body makes two successive compensable movements, then the resulting global movement is also compensable. In other words, this class of movements has the mathematical structure of a group, whose identity element is stationarity. We shall call an element of this group a *compensable transformation* of the exteroceptive body. In the same way, we can define the group of compensable transformation of the environment and the group of compensated transformations of the body-environment system. By their definitions, these three groups are very closely related to each other.
- Compensability implies that there is something in common between certain body movements and certain environmental movements. It

now (but not before) becomes possible to say that body and environment are immersed in a single entity that we call *space*. Note that it is through the use of a sensorimotor approach that we have attained the notion of space: there can be no notion of compensability when we make only passive observations. It is through actions that arises the idea of an environment distinct from the body and through actions that is born the idea of a common medium through which to describe the body and the environment. In a similar vein, Poincaré (1895, 1902) pointed out the radical incommensurability between sensations originating in different sensory modalities, unified only by common motor acts needed to reach their sources.

Certain compensated transformations have an additional, special structure. They form a non-trivial subgroup, in the sense that the sensory consequences of two successive transformations of that particular kind may depend on the order in which they have been performed. Other compensated transformations can be applied before or after any other compensated transformations (including the previously mentioned ones) without changing anything from the sensory point of view. We will call the first transformations *rigid transformations* and the others *attribute transformations*. Indeed, the first group of transformations defines space, and the other transformations define changes over nonspatial attributes.

The terminology of *rigid transformations* comes from the fact that these transformations are exactly related to those spatial transformations of the body-environment system leaving the sensory inputs fixed. Thus, any metric defined over representations of space based on the sensory inputs only will have to be constant over these latter changes.

Since our organism is assumed to be totally ignorant about itself and its surroundings, it has no choice but to define rigidity; the same problem was encountered for the notion of stationarity. This means that if, for example, the physical laws obeyed by light changed according to the positions of the sensors, there would be a difference with the Euclidean notion of rigidity. But physicists themselves encounter the same problem in their conception of space, since they must also assume as given the notion of a rigid measuring rod.

We have attained the nontrivial conclusion that without observing its exteroceptive body “from the outside,” the brain can deduce that there exist transformations of the body-environment system that its exteroception is not sensitive to. These compensated transformations provide the brain with what it needs to ground its understanding of geometry.

It is worthwhile to point out the relation of our approach with that of Droulez and Berthoz (1991), Blakemore, Frith, and Wolpert (2001), and Blakemore, Wolpert, and Frith (2002), where it is proposed that the cerebellum attempts to subtract a prediction of the sensory consequence of its movements from the actually observed consequences in order to deduce

the changes that must have occurred in the environment. The similarity lies in the very general idea of analyzing the sensorial consequences of a movement of the body alone to understand the changes of the environment. This is a classical idea today. But it usually relies on a kind of platonic a priori about the existence of space and assumes that the role of the brain is to map its sensory inputs to some kind of objective archetype of the world and try to understand its sensations in relation to this abstract world. Here we used terms such as “representations of the state of the exteroceptive body” to describe what *we* (or the brain) conceive this world to be, without any relationship with an a priori model.

3 Mathematical Sketch

In order to make the preceding discussion more precise and in order to derive a simple, neuronally plausible algorithm, we present a sketch of a mathematical formalization. We illustrate only the essential aspects of our approach to show in a few steps how it is possible to deduce the dimension of the manifold of rigid transformations of outside “physical” space. The appendix provides suggestions for a more realistic implementation, and a second article will show, beyond the discovery of dimensions, how the group structure of these rigid transformations can be accessed and used.

We think that the problem we want to answer is precisely the problem addressed in differential geometry. Indeed, a usual way of introducing the aim of differential geometry is transparently summarized by saying that “to consider \mathcal{S} as a manifold means that one is interested in investigating those properties of \mathcal{S} that are invariant under coordinate transformations” (Amari & Nagaoka, 1993). If we think of the sensorimotor system as a highly redundant parameterization system to move in this manifold and make observations on it, then the goals are strikingly identical. The structure of the world consists in those properties that will be imposed on any sensorimotor system; the rest is an artifact of the specificities of bodies. It is in this spirit that we use the language of differential geometry, with the conviction it can provide insights.

Consider an environment whose set of all states E is a manifold \mathcal{E} of dimension e . Suppose the set of all observed sensory inputs S is a manifold \mathcal{S} of dimension s , and the set of all possible outputs M is a manifold \mathcal{M} of dimension m . Finally, suppose the environment imposes a “smooth enough” (meaning that we will consider the problem only in a region excluding the singularities of ψ) relation between sensory signals S and motor commands M :

$$S = \psi(M, E). \quad (3.1)$$

Note that with our definitions, we have $S = \psi(\mathcal{M} \times \mathcal{E})$.

Although the mathematics does not require it, to facilitate the discussion below, we shall consider the case where the manifolds involved are embedded in finite vector spaces and that (S, M) is a vector whose coordinates reflect the neural activities of the sensory and motor neurons at some time t . Making this choice has the consequence that the sensory inputs will have to be determined instantaneously by the motor output and the environment states. This is thus inappropriate for cases where the sensory inputs are determined by the integration over time of motor commands, as is the case when we control the rotation speed of the wheels of a robot, for instance. Our example would apply instead to a case where the neural motor activity would be enough to infer muscle elongation and thus the position of the body. However, we wish to emphasize that this choice of instantaneous relation between sensory and motor activities is not necessary from the mathematics and is mainly used to make the presentation here more intuitive. Furthermore, it will allow us to identify proprioception very easily.

Indeed, because we argue that geometry arises through compensability and that proprioceptive inputs cannot be compensated by any change of the environment, the first thing we want to do is to distinguish proprioceptive from exteroceptive inputs. This can easily be done in the previous model by locating inputs that do not change when a given definite motor order is issued and maintained. From now on, we shall be concerned only with exteroceptive inputs.

Following the method frequently adopted in motor control (Atkeson & Schaal, 1995; Vijayakumar & Schaal, 2000; Baraduc, Guigon, & Burnod, 2001, although we are not using a motor control approach $M = \varphi(S)$, but an approach on the contrary based on the observation of sensory consequences of motor commands) and the standard mathematical approach for dealing with manifolds, we shall focus on the tangent space $\{dS\}$ of S at some point $S_0 = \psi(M_0, E_0)$.

This is allowed by the smoothness of ψ , and it is important to note that this provides an exact theoretical ground for the considerations below, while the robustness of the local linear approximation we will make in practice is a different issue. Our aim was to show that the dimension of the rigid group of space is accessible through sensory inputs and to demonstrate the conceptual framework needed to access it. The actual implementation we used had no other purpose than illustration, and the question of its robustness for practical robotic applications has no incidence on our more theoretical point. The simple, neuronally plausible (Sejnowski, 1977; Oja, 1982; Sanger, 1989) tool we used to estimate these dimensions is principal component analysis (PCA), but any other method of dimension estimation would have been suitable.

This being said, two natural subspaces can be identified in $\{dS\}$: the vector subspace $\{dS\}_{dE=0}$ of sensory input variations due to a motor change only and the vector subspace $\{dS\}_{dM=0}$ of sensory input variations due to an

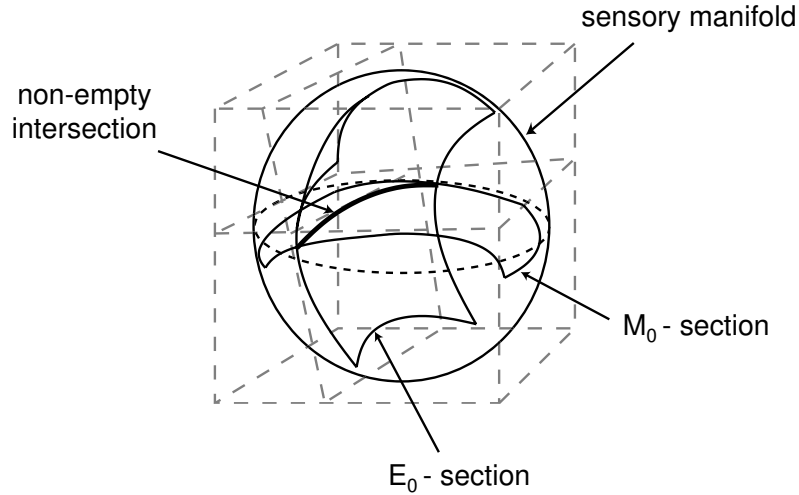


Figure 2: The sensory manifold in the neighborhood of S_0 , the E_0 and M_0 -sections (see text). These two manifolds are transverse, and their intersection is the manifold of the sensory inputs accessible through either motion of the exteroceptive body or motion of the environment.

environment change only. Since we have

$$dS = \frac{\partial \psi}{\partial M}|_{(M_0, E_0)} \cdot dM + \frac{\partial \psi}{\partial E}|_{(M_0, E_0)} \cdot dE, \quad (3.2)$$

we can remark that

$$\{dS\} = \{dS\}_{dM=0} + \{dS\}_{dE=0}. \quad (3.3)$$

What are the manifolds to which these subspaces are tangent? Starting from (M_0, E_0) , we can consider the sensory inputs obtained through variations of M only (E_0 -section $\equiv \psi(E_0, \mathcal{M})$) and sensory inputs obtained through variations of E only (M_0 -section $\equiv \psi(\mathcal{E}, M_0)$) (see Figure 2). $\{dS\}_{dE=0}$ and $\{dS\}_{dM=0}$ are the tangent spaces at the point S_0 of these manifolds, and the fact that their vectorial sum is the overall tangent space of \mathcal{S} means, by the definition of transversality, that the two sections are transverse (see, e.g., Laudenbach, 1999, for a basic description of transversality in submanifolds of \mathcal{R}^n). We will call $\mathcal{C}(M_0, E_0)$ their intersection, which is thus a manifold as well.

When the body is stationary, the dimension $\{dS\}_{dM=0}$ gives the number e of variables necessary for a local description of the environment. When the environment is stationary, the dimension $\{dS\}_{dE=0}$ gives the number $p \leq m$

of variables necessary to explain the variations in exteroceptive signals due to body motions, that is, the number of variables describing the exteroceptive body. When both M and E vary, $\{dS\}$ is not of dimension $p + e$ because the vector spaces generated by $\frac{\partial \psi}{\partial M}$ and $\frac{\partial \psi}{\partial E}$ do not necessarily have null intersection. Certain exteroceptive changes can be obtained equally from either dE or dM , as is the case when we move along $\mathcal{C}(M_0, E_0)$. This reduces the dimensionality of dS compared to the case where it is possible to identify the origin of the change unambiguously. When sensory changes from dE and dM mutually compensate for one another, we shall say that there has been a compensated infinitesimal movement of the body-environment system, and we shall say that the corresponding infinitesimal movement of the body (or the environment) is compensable (thus, implicitly, the term of compensated movement means a change of body-environment, while compensable movement means a change of either one alone). For simplicity we will drop the word *infinitesimal*, but it should always be kept in mind since we will mostly be dealing with the tangent spaces. This is justified because it is a basic result that the dimension of a tangent space is the dimension of the underlying manifold.

Since the exteroceptive body representation and the environment representation imply parameterizations of $\{dS\}_{dM=0}$ and $\{dS\}_{dE=0}$, there is a natural one-to-one correspondence between the set of compensated movements and the set of compensable movements of the body. The basic idea is that given a compensable movement of the body, only the environment movement yielding the opposite sensory consequence will, when taken together, yield a compensated movement of the body-environment system.³

We will now search for the relationships between the dimensions of all the entities we have defined. We have shown that the dimension of the space of compensated movements is the same as the dimension of the space of compensable body movements, and this dimension is accessible because there is a trivial one-to-one mapping of this space to $T = \{dS\}_{dM=0} \cap \{dS\}_{dE=0}$. Indeed, that $dS_{dM=0}$ is compensable means:

$$\begin{aligned} \exists dS_{dE=0} \text{ such that } dS_{dM=0} + dS_{dE=0} = 0 &\Leftrightarrow dS_{dM=0} \\ &= -dS_{dE=0} \Leftrightarrow dS_{dM=0} \in T. \end{aligned}$$

³ This is enough for our purpose, but (for the extension of this work) it is important to understand that this can be extended to noninfinitesimal movements by integration: given a compensable movement of the body (i.e., a curve tangentially compensable at any time), we can construct the movement of the environment whose tangent sensory change will cancel at any time the tangent sensory change yielded by the compensable movement. This is symmetrically true for compensable movements of the environment.

Thus, the space of the compensated movements has the dimension of T . But since

$$\begin{aligned} \dim\{dS\}_{dM=0} + \{dS\}_{dE=0} &= \dim\{dS\}_{dM=0} \\ &+ \dim\{dS\}_{dE=0} - \dim\{dS\}_{dM=0} \cap \{dS\}_{dE=0}, \end{aligned}$$

we finally have, with the use of equation 3.3,

$$d = p + e - b, \quad (3.4)$$

where $b = \dim\{dS\}$ and $d = \dim T = \dim \mathcal{C}(M_0, E_0)$.

We have consequently deduced the dimension of the manifold of the compensated transformations of the world. Note that this is actually more fundamental than simply the number of variables needed to describe this world: the manifold of compensable transformations is the operational aspect of the relation to the world that an organism will be most interested in.

In a subsequent, more technical article (in preparation), we demonstrate how the group structure of the underlying transformations can be accessed and defined through $\mathcal{C}(M_0, E_0)$. This will provide a way for the naive organism to understand its body and produce at will particular rigid transformations of its exteroceptive body, such as translations or rotations for instance. The dimension of the rigid group and the dimension of space will be found through the study of these transformations. For now, if we make the additional assumption that the organism can choose to perform only spatial movements, then we can say that we have accessed the dimension of the rigid group defining the space that the organism is embedded in.

A last point must be made concerning the distinction between proprioceptive and exteroceptive sensory inputs. In the preceding mathematical discussion, we have assumed we are dealing only with exteroception and that proprioception has been set aside. It might be thought that this is an unnecessary step, since in equation 3.4, proprioceptive dimensions surely would cancel out, since they would contribute equally to p and b . However, this is in general false. If we consider a case where the representation of the exteroceptive body intersects with the representation of the proprioceptive body, then describing this intersection will, in a moving environment, require two times the set of variables required in a fixed environment. Indeed, in a moving environment, exteroception and proprioception will be unlinked.

4 Experiments

4.1 Method. We present the results of a simulation of the example of the articulated arm described above. We additionally describe simulations for two further cases with interesting modifications of the organism. The details of the three experiments can be found in the appendix. A summary

Table 1: Summary of the Three Experiments.

Characteristics	Organism 1	Organism 2	Organism 3
Dimensions of motor commands	40	100	100
Dimensions of exteroceptive inputs	40	80	80
Number of eyes	2	4	4
Diaphragms	None	Reflex	Controlled
Number of lights	3	5	5
Light luminance	Fixed	Variable	Variable
Dimensions found for body (p)	12	24	28
Dimensions found for environment (e)	9	20	20
Dimensions found for both (b)	15	38	41
Deduced dimension of rigid group (d)	6	6	7

Notes: Proprioception does not play a role in the calculation and so is not shown in the table. The estimations given here are obtained from Figures 3c and 4. In Organism 3, the group of compensated transformations is different from the orthogonal Euclidean group because the organism has control over a nonspatial aspect of its body, namely the diaphragm aperture.

of the results is presented in Table 1. It should be stressed that the same kind of simulation could be done for any other arbitrary kind of device with sensory inputs and motor outputs.

In the first experiment, the arm had four joints and two eyes, and the environment consisted of three lights. Each eye consisted of a composite “retina” with 20 omnidirectionally (i.e., not directional) sensitive photo sensors mounted rigidly on a small, flat surface, attached to the end of a “finger,” one for each eye.

Each joint had four proprioceptive sensors whose output depended on the position of the joint, according to a fixed, randomly assigned law. The orientation of the eyes provided no proprioception.

The motor command moving the device was a 40-dimensional vector, which was converted by a fixed random function to the 12 values that determined the 3D spatial coordinates of the surfaces holding the two eyes and their orientations.

These particular choices were arbitrary: the purpose was merely to simulate a complicated sensorimotor relation that was unknown to the brain and had the property that the number of dimensions of the motor commands and of the sensory inputs should be high compared to the number of degrees of freedom of the physical system.

In the second experiment, we considered a more complex device with an arm having 10 joints, bearing 4 eyes. Each eye had a diaphragm or attenuator with an automatic “pupil reflex” that reduced light input to it in such a way that total illumination for the eye was constant. There were five light sources in the environment, and we now allowed their inten-

sity to vary. The dimensionality of the motor command and sensory input was also increased, respectively, to 100 (determining the 24 degrees of freedom of the four eyes, each having three positional and three orientational degrees of freedom) and 120 (determined by the 20 photosensors on each of the four eyes plus 40 proprioceptors). Again the purpose was to show that the complexity of the sensorimotor coupling was not a theoretical obstacle to our approach; neither were nonspatial body changes like the pupil reflex or nonspatial changes in the environment, like light intensity.

The third experiment was identical to the second except that now we made what we shall see is a theoretically very important modification: the diaphragms were now controlled by the organism instead of being determined automatically by the total illumination.

To do the simulation, we went through the four-stage procedure described in the previous section:

1. Proprioceptive input was separated from exteroceptive input by noting that proprioceptive input remains silent when no motor commands are given, whereas exteroceptive input changes because of environmental change.
2. We estimated the number of parameters needed to describe the variation in the exteroceptive inputs when only the environment changes. The algorithm issues no motor commands and simply calculates the covariance matrix of the observed environment-induced variations in sensory inputs. The dimension estimation is done by considering the eigenvalues of this covariance matrix. The eigenvalues λ_i should fall into two classes: a class with values all equal to zero and a class with nonzero values. We separated the two classes by a clustering method (see the appendix). The number of nonzero eigenvalues was taken as the number of dimensions.
3. We estimated the number of parameters needed to describe the variation in the exteroceptive inputs when only the body moved. The environment is kept fixed, and the algorithm gives random motor commands (see the appendix). We observe the covariance matrix of the resulting changes and estimate the dimension from the number of nonzero eigenvalues in the same way as before.
4. We estimate the number of parameters needed to describe the changes in exteroceptive inputs when both the body and the environment change. The environment is changed at random, and the organism gives random motor commands. The number of nonzero eigenvalues of the covariance matrix is obtained as before.

4.2 Results for Organism 1. We see that despite the high dimensionality of the outputs, the exteroceptive, and the proprioceptive inputs (40, 40, 20,

respectively), the brain is able to:

- Distinguish the 20 proprioceptive and 40 exteroceptive sensor inputs (see Figure 3a).
- Determine that only 12 parameters are necessary to represent its exteroceptive body—this corresponds to the three position and three orientation dimensions for each of the two eyes (see Figures 3b and 3c).
- Determine that 9 parameters are necessary to describe the environment, corresponding to three spatial dimensions for each of the three lights (see Figures 3b and 3c).

These results would seem to lead to the conclusion that the brain should observe $12 + 9 = 21$ degrees of freedom in the exteroceptive inputs when it lets body and environment move simultaneously, but instead only 15 are found (see Figures 3b and 3c). The brain thus concludes that there is a group of compensated movements of dimension $21 - 15 = 6$, which we know to be the Lie group of orthogonal transformations (3 translation and 3 rotations).⁴

4.3 Results for Organisms 2 and 3. In the second experiment, the algorithm deduced that 24 variables were needed to characterize the exteroceptive body, 20 were needed for the environment, and 38 were needed to characterize the exteroceptive inputs when both environment and body moved simultaneously (see the appendix). The group of compensated movements constituted by the brain therefore remained coherent with the usual Euclidean group, since the brain again arrived at a group of dimension $20 + 24 - 38 = 6$ (which is the same Lie group as before). We see that the notion of space constituted by the brain is insensitive to major variations in the complexity of the physical structure of the device and the details of its sensorimotor coupling.

The third experiment involved the same organism as the preceding example, but the aperture of the diaphragms was now controlled by the brain instead of being determined by an automatic reflex. This time the brain determined that 28 variables were needed to characterize the exteroceptive body. Twenty were needed for the environment and 41 for the exteroceptive inputs when both environment and body moved. Luminance in this example is a compensable variable, and we indeed found that we now have $28 + 20 - 41 = 7$ compensated variables instead of 6.

⁴ An elementary presentation of Lie groups can be found in Burke (1985). The orthogonal transformations are clearly a subset of the set of compensated movements, and the equality of dimensions guarantees equality of the groups. The idea that sensory information should be invariant under the Lie group of orthogonal transformations is a useful concept in artificial vision (Van Gool, Moons, Pauwels, & Oosterlinck, 1995) and image analysis (Rao & Ruderman, 1999).

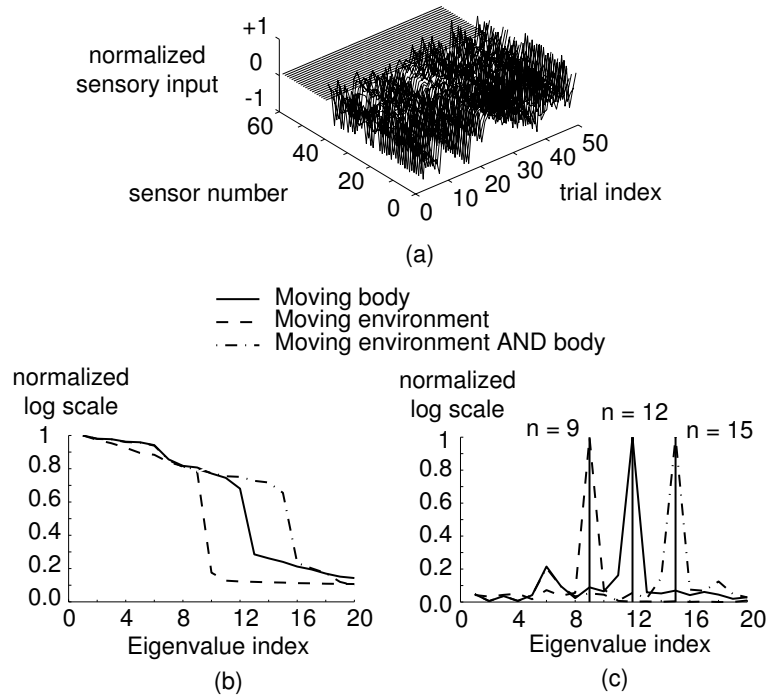


Figure 3: (a) Changes in the different inputs when no motor commands are given and the environment changes. The first 40 sensors depend on the environment (exteroceptive) and the 20 last ones do not (proprioceptive). (b) Eigenvalues of the covariance matrices in the three cases described in the text (normalized log scale : logarithm of the eigenvalues renormalized to lie in the range $[0,1]$). (c) Ratio of eigenvalue i to eigenvalue $i + 1$, where i is the eigenvalue index. The dimensions of the tangent spaces are taken to be equal to the number of significantly nonzero eigenvalues, and the maximum of this ratio indicates the biggest change in the order of magnitude of the eigenvalues (see the appendix).

The group structure, and more precisely commutativity of the transformation involving only luminance variations with any other transformation, still theoretically allows for the distinction between spatial transformations and attribute changes. But it is worthwhile to recall that Poincaré thought of geometry in terms of voluntary body motions. It would seem that he considered that only spatial changes were subject to voluntary control. If space is thus defined by the voluntary accessible set of sensory inputs, then the control of our diaphragms would surely have led us to a different conception of space than the one we have. If we define space through the properties of

commutativity of transformations operating on it, which is the hypothesis we favor, then the diaphragms are of no importance.

A second aspect of this third experiment should not be dismissed; it concerns the notion of environmental attributes. By allowing luminance to vary independently in the five lights, one might have expected that the brain would find five additional variables as compared to the case where luminance is fixed. However, the algorithm deduces that only a single additional compensable dimension need be introduced. This is because compensability of luminance with motions of the body-environment configuration is possible only in the case where all five lights vary simultaneously in luminance by a common factor. Thus, we see that the algorithm has extracted the presence of an additional feature dimension, luminance, independent of the number of lights that are present.

We wish to stress for the last time that the simulations presented here are considered simply as illustrations of our approach. We claim nothing about the optimality or robustness of the methods we have used. There is an active literature on the practical problems involved in estimating dimensionality and, more interesting, the parameterization of manifolds (Tenenbaum, 1998; Roweis & Saul, 2000; Tenenbaum, Silva, & Langford, 2000; Verbeek, Vlassis, & Kröse, 2002). Our purpose here was to show how, when taken in conjunction with a sensorimotor rather than merely sensory approach, such tools could be used by an organism to generate the notion of space.

5 Conclusion

We have shown that thanks to a sensorimotor approach, the notion of rigid group of space, where space is an abstract frame in which the brain has good reasons to embed its body and its environment, is accessible to an organism without any a priori knowledge. This shows why, in robotics and neurobiology, it may be fundamental to take sensory and motor information into account simultaneously. Indeed, we have shown that doing so provides a meaningful way of reducing the dimensionality of the problem of interacting with the environment. This is directly of interest for roboticians concerned with unsupervised learning. It is indirectly of interest for neurobiologists and cognitive scientists, since it favors the idea suggested in O'Regan & Noë (2001) that the brain should be studied from a sensorimotor approach rather than a stimulus-based or motor control approach: until now, it has never been shown how such an approach could access the notion of space, yet it is a notion we all know to be present in our brains.

Our approach represents a basically different approach than the dimension-reduction techniques that are usually employed, for instance, in pattern classification tasks for passive observers. Consider, for example, Roweis and Saul (2000), which does dimension reduction on a set of faces with different expressions. In this situation, the variables that can be used to parameter-

ize facial expressions will have no simple relation to the dimension of the space in which the faces are embedded. If such a technique is applied to samples of faces that can be both rigidly rotated and can change expression, the parameters determining these two types of change would not be differentiated. On the other hand, by the use we are suggesting here of an active observer and the notion of compensated variables, it is possible for our algorithm to make a principled distinction between rigid deformations like rotations and nonrigid variations like facial expression changes. Our approach induces a difference between simple feature spaces that do not distinguish attributes and geometry, and structured geometrical spaces.

Finally, perhaps the main interest of our approach is philosophical. Spatial reasoning is the basis of our comprehension of the world and our abstraction abilities, so much so that the nature of space itself has been considered to preclude scientific investigation. Ever since the revolution of non-Euclidean geometry more than a century ago, our Euclidean intuition of space has been considered to derive from our experience with the world and not from a mental *a priori*. But then if our notions of geometry derive from our relation with the world, to what extent do these notions depend on the world and to what extent do they depend on the way our sensorimotor systems are constructed? Could it be that the intuition of three-dimensional Euclidean space is a consequence of the structure of humans' sensorimotor systems rather than corresponding to "real" properties of the outside world?

To answer this question, we would have to know what we mean by "real" properties of the world, even though we obtain information about it only through our bodies and our sensors. This is problematic, since to understand the world, we must know about how our bodies and sensors are constructed. But how can we formulate such knowledge without presupposing a world within which the body and its sensors are embedded and whose concepts we use to describe those very bodies and sensors? It is because of this fundamental problem that we think it is fruitful to develop an abstract approach such as ours, in which the details of the workings of sensors and effectors play no role.

Our approach has shown that one reasonable deduction that brains can derive from the sensorimotor constraints they are exposed to is the notion of compensated movements. A brain can naturally infer that the structure of the derived compensated transformations of the world will be coherent with transformations derived by any other brain with the same sensorimotor capabilities but different details of implementation. Indeed, since the compensated group is precisely those changes of the body-environment system leaving perception unchanged, if there is a one-to-one smooth mapping from the perception of one organism to the perception of the other, then they will end up with the same rigid group. The subsequent consensual universality of this group makes it a good candidate for the status of physical reality. We believe that it may be this which lies at the basis of our notion of the orthogonal group of geometrical space.

Appendix

A.1 Experimental Details. The motor commands were simulated in the following way:

$$\begin{aligned} (Q, P, a) &= \sigma(W_1 \cdot \sigma(W_2 \cdot M - \mu_2) - \mu_1) \\ L &= \sigma(V_1 \cdot \sigma(V_2 \cdot E - \nu_2) - \nu_1) \\ S_{i,k}^e &= d_i \sum_j \frac{\theta_j}{\|P_i + \text{Rot}(a_i^\theta, a_i^\varphi, a_i^\psi) \cdot C_{i,k} - L_j\|^2} \\ S_i^p &= \sigma(U_1 \cdot \sigma(U_2 \cdot M - \tau_2) - \tau_1), \end{aligned}$$

where $W_1, W_2, V_1, V_2, U_1, U_2$ are matrices with coefficients drawn randomly from a uniform distribution between -1 and 1 , as are also the vectors $\mu_1, \mu_2, \nu_1, \nu_2, \tau_1, \tau_2$. This is equivalent to the choice of the measurement unit for the signals of our model. σ is an arbitrary nonlinearity, here the hyperbolic tangent function. The $C_{i,k}$ are drawn from a centered normal distribution whose variance, which can be understood as the size of the retina, was so that the sensory changes resulting from a rotation of the eye were of the same order of magnitude as the ones resulting from a translation of the eye.

Here are the notations used in the definition of the simulated sensorimotor law:

$Q = (Q_1, \dots, Q_q)$	positions of the joints
$P = (P_1, \dots, P_p)$	positions of the eyes
$a_i^\theta, a_i^\varphi, a_i^\psi$	Euler angles for the orientation of eye i
$\text{Rot}(a_i^\theta, a_i^\varphi, a_i^\psi)$	rotation matrix for eye i
$C_{i,k}$	relative position of photosensor k within eye i
$d = (d_1, \dots, d_q)$	apertures of diaphragms
$L = (L_1, \dots, L_p)$	positions of the lights
$\theta = (\theta_1, \dots, \theta_p)$	luminances of the lights
$S_{i,k}^e$	sensory input from exteroceptive sensor k of eye i
S_i^p	sensory input from proprioceptive sensor i
M, E	motor command and environmental control vector

In Organism 1, motor command M is a vector of size 40 and S a vector of size 40. The organism has four joints and two eyes with 20 photosensors each. The eyes are free to rotate in any way, including through torsion movements. θ and d are constants drawn at random in the interval $[0.5, 1]$.

Figures 3b and 3c present the results obtained for a linear approximation in a neighborhood of size 10^{-8} (this is to be understood with respect to the unit measure given by the previous matrix coefficients). Sensory inputs were generated from 50 motor commands and 50 environmental positions following a normal distribution with mean zero and standard deviation 10^{-8} . Coordinates differing from zero by more than the standard deviation were put equal to zero. This neighborhood size is the one that yielded the most significant separation between very small eigenvalues and non-null ones, but results could be found from about 10^{-5} . Generally, the size of validity for the linear approximation is related to the curvature of the sensory manifold, and there is no general answer to the question of what size the neighborhood should have.

Out of these 50 motor commands $\{dM_i\}$ and 50 environmental positions $\{dE_j\}$, we produced 50 sensory input changes $\{dS_i\}_{dE=0}$ resulting from the world configurations $\{(M_0 + dM_i, E_0)\}$, 50 sensory input changes $\{dS_j\}_{dM=0}$ resulting from the world configurations $\{(M_0, E_0 + dE_j)\}$, and $50 \times 50 = 2500$ sensory input changes $\{dS_{i,j}\}$ resulting from the world configurations $\{(M_0 + dM_i, E_0 + dE_j)\}$.

The results of the other two experiments (see Figure 4) were obtained with environmental lights having variable luminance. The organism had 10 joints and a diaphragm for each of its four eyes. Motor commands were of dimension 100. For Organism 2, the diaphragm aperture d_i of eye i was defined by the equation

$$\sum_k S_{k,i} = 1,$$

that is, the aperture adjusted automatically so as to ensure constant total illumination for eye i . For Organism 3, d was under its own control:

$$(Q, P, a, d) = \sigma(W_1 \cdot \sigma(W_2 \cdot M - \mu_2) - \mu_1).$$

Theoretically the eigenvalues of the covariance matrices of the three sample sets $\{dS_i\}_{dE=0}$, $\{dS_j\}_{dM=0}$, and $\{dS_{i,j}\}$, should fall into two classes—one with zero values and one with nonzero values. To distinguish these two classes we used a clustering method and supposed that the two classes V_1 and V_2 were such that each λ of V_1 was more comparable in size to other λ 's of V_1 than to all those of V_2 , and conversely. Finding the boundary between the two classes can thus be done by ordering the λ_i in decreasing order, and locating the value of i such that the ratio between λ_i and λ_{i+1} is largest (see Figures 3c and 4). We could also have used an approach similar to Minka (2001). It should be noticed that the nullity or non-nullity of the eigenvalues reflects a characteristic of space, but their absolute value reflects a characteristic of the sensorimotor system. This is where lies the practical aspect of our problem as compared with the theoretical one.

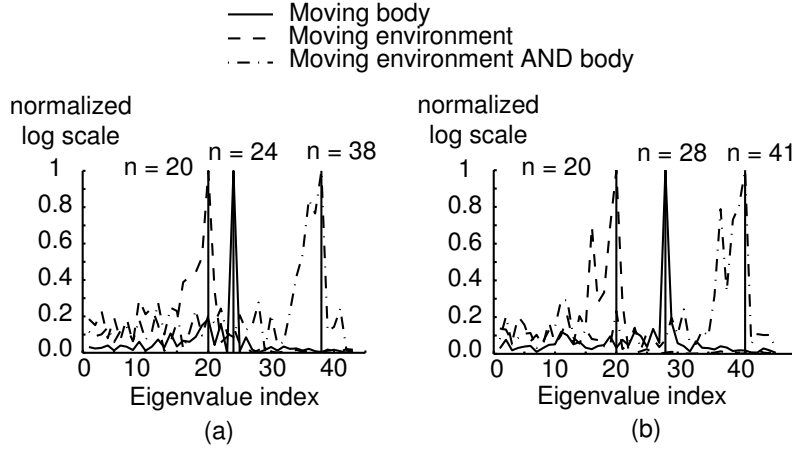


Figure 4: Dimension estimation when the diaphragm aperture is determined by reflex (a) and controlled by the organism (b).

A.2 Extensions of the Model. The concrete examples we have presented here might give the impression that the approach is limited to cases where motor commands and sensory input are defined as instantaneous vectors. This would be unrealistic, since motor commands and neural afference are usually extended in time. But this poses no problem for our approach, since \mathcal{M} , \mathcal{E} , and \mathcal{S} need only to have a manifold structure, not a vectorial structure. For simplicity, we chose vectors representing the motor commands and sensory inputs at a particular time t , but in general, M and S could be vector functions, and the sensorimotor relation would then be an equation of functionals. In the finite case, our approach could be applied directly, and performing the PCA would then require making a local approximation of this functional relation in a manner analogous to that proposed by Fod, Matarié, and Jenkins (2000) for motor control. In the nonfinite case, we think there should also be a way to use our method, since $\mathcal{C}(M_0, E_0)$ is finite even if \mathcal{S} , \mathcal{M} , and \mathcal{E} are not.

When the functional equation is given as a differential equation, the sensorimotor relation remains instantaneous. We will in general be able to write it in implicit form,

$$\psi(S(t), \dot{S}(t), \dots, M(t), \dot{M}(t), \dots, E(t)) = 0, \quad (\text{A.1})$$

and then determine the variables that are functions of $S(t)$, $\dot{S}(t)$, \dots , which can be accounted for in terms of $M(t)$, $\dot{M}(t)$, \dots .⁵ From there on, we can apply the same reasoning we used for the exteroceptive variables described in this article. This extension is one of the objects of our current research.

Acknowledgments

We thank Olivier Coenen for his support and remarks and Jim Clark, Jacques Droulez, Daniel Bennequin, Alexandre Pouget, Jean Petitot, Patrick Gros, and Christoph Zetzsche for their enlightening observations. This work was partly funded by the European Project IST-2001-35271.

References

- Amari, S.-I., & Nagaoka, H. (1993). *Methods of information geometry*. New York: Oxford University Press.
- Atkeson, C. G. & Schaal, S. (1995). Memory-based neural networks for robot learning. *Neurocomputing*, 9, 243–269.
- Baraduc, P., Guigon, E., & Burnod, Y. (2001). Recoding arm position to learn visuomotor transformations. *Cerebral Cortex*, 11(10), 906–917.
- Blakemore, S. J., Frith, C. D., & Wolpert, D. M. (2001). The cerebellum is involved in predicting the sensory consequences of action. *Neuroreport*, 12(9), 1879–1884.
- Blakemore, S. J., Wolpert, D. M., & Frith, C. D. (2002). Abnormalities in the awareness of action. *Trends. Cogn. Sci.*, 6(6), 237–242.
- Burke, W. L. (1985). *Applied differential geometry*. Cambridge: Cambridge University Press.
- Droulez, J., & Berthoz, A. (1991). A neural network model of sensoritopic maps with predictive short-term memory properties. *Proceedings of the National Academy of Sciences*, 88, 9653–9657.
- Fod, A., Matarié, J., & Jenkins, O. C. (2000). Automated derivation of primitives for movement classification. Paper presented at the International Conference on Humanoid Robotics.
- Kandel, E. R., Schwartz, J., & Jessell, T. M. (2000). *Principles of neural science*. New York: McGraw-Hill.
- Laudenbach, F. (1999). *Calcul différentiel et intégral*. Paris: Ecole Polytechnique.
- Minka, T. P. (2001). Automatic choice of dimensionality for PCA. In T. K. Leen, T. G. Dietterich, & K.-R. Müller (Eds.), *Advances in neural information processing systems*, 13, (pp. 598–604). Cambridge, MA: MIT Press.
- Oja, E. (1982). A simplified neuron model as a principal component analyser. *Journal of Mathematical Biology*, 15, 267–273.

⁵ Linear combinations, in the case of a local linearization, which could, again, be determined by a PCA. Such a linearization could perhaps also be applied globally, since linear differential equations can be used to explain a wide range of physical problems, and in particular probably those that can be apprehended quickly by our brains.

- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 883–917.
- Poincaré, H. (1895). L'espace et la géométrie. *Revue de Métaphysique et de Morale*, 3, 630–646.
- Poincaré, H. (1902). *La science et l'hypothèse*. Paris: Flammarion.
- Rao, R. P. N., & Ruderman, D. L. (1999). Learning Lie groups for invariant visual perception. In M. S. Kearns, S. Solla, & D. Cohn (Eds.), *Advances in neural information processing systems*, 11, (pp. 810–816). Cambridge, MA: MIT Press.
- Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290, 2323–2326.
- Sanger, T. (1989). *Optimal unsupervised learning in feedforward neural networks* (Tech. Rep.). Cambridge, MA: AI Lab, MIT.
- Sejnowski, T. (1977). Storing covariance with nonlinearly interacting neurons. *Journal of Mathematical Biology*, 4, 303–321.
- Tenenbaum, J. B. (1998). Mapping a manifold of perceptual observations. In M. Kearns, M. Jordan, & S. Solla (Eds.), *Advances in neural information processing systems*, 10. Cambridge, MA: MIT Press.
- Tenenbaum, J. B., Silva, V. de, & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290, 2319–2323.
- Van Gool, L., Moons, T., Pauwels, E., & Oosterlinck, A. (1995). Vision and Lie's approach to invariance. *Image and Vision Computing*, 13(4), 259–277.
- Verbeek, J. J., Vlassis, N., & Kröse, B. (2002). *Coordinating mixtures of probabilistic principal component analyzers* (Tech. Rep. IAS-UVA-02-01) Amsterdam: Computer Science Institute, University of Amsterdam.
- Vijayakumar, S., & Schaal, S. (2000). Fast and efficient incremental learning for high-dimensional motion systems. *International Conference on Robotics and Automation (ICRA2000)* (pp. 1894–1899).

6.2 Article : Perception of the structure ...

L'article "*Perception of the structure of the physical world using unknown multi-modal sensors and effectors*" est la suite relativement directe de ce qui précède, montrant que, sous certaines conditions plus ou moins réalistes, la variété des dépendances sensorimotrices d'un système caractérise non seulement la dimension du groupe de symétries de l'espace sous-jacent, mais également les différents types de mouvements dans cet espace : translations, et ensembles de rotations autour d'un même point. En d'autres termes, alors que l'article précédent montrait que des dépendances sensorimotrices associées à des groupes de symétries de dimensions différentes sont distinctes, celui-là montre que les mouvements du dispositif associés à des générateurs différents du groupe de symétrie sont distincts.

Avec le recul, le cadre général proposé par cet article peut paraître artificiellement compliqué, en particulier par rapport à la reformulation suggérée en section 6.5, et même criticable sur plusieurs points. Il ne s'agit donc pas de s'attacher au détail de l'argumentation présentée dans l'article, mais simplement à son idée principale : 1) un même mouvement du dispositif sensoriel dans les différentes configurations de l'environnement permet de définir (modulo des hypothèses avantageuses) un champ de vecteurs sur la variété sensorielle, 2) l'ensemble des champs de vecteurs obtenus de cette manière ne sont pas équivalents les uns aux autres vis-à-vis du crochet de Lie, 3) il en découle que les mouvements du dispositif sensoriel (et donc, de manière symétrique, de l'environnement), ne sont pas équivalents les uns aux autres.

Perception of the structure of the physical world using unknown multimodal sensors and effectors

D. Philipona

Sony CSL, 6 rue Amyot
75005 Paris, France

david.philipona@polytechnique.org

J.K. O'Regan

Laboratoire de Psychologie Expérimentale, CNRS
Université René Descartes, 71, avenue Edouard Vaillant
92774 Boulogne-Billancourt Cedex, France

<http://nivea.psych.univ-paris5.fr>

J.-P. Nadal

Laboratoire de Physique Statistique, ENS
rue Lhomond
75231 Paris Cedex 05

O. J.-M. D. Coenen

Sony CSL, 6 rue Amyot
75005 Paris, France

Abstract

Is there a way for an algorithm linked to an unknown body to infer by itself information about this body and the world it is in? Taking the case of space for example, is there a way for this algorithm to realize that its body is in a three dimensional world? Is it possible for this algorithm to discover how to move in a straight line? And more basically: do these questions make any sense at all given that the algorithm only has access to the very high-dimensional data consisting of its sensory inputs and motor outputs?

We demonstrate in this article how these questions can be given a positive answer. We show that it is possible to make an algorithm that, by analyzing the law that links its motor outputs to its sensory inputs, discovers information about the structure of the world regardless of the devices constituting the body it is linked to. We present results from simulations demonstrating a way to issue motor orders resulting in “fundamental” movements of the body as regards the structure of the physical world.

1 Introduction

What is it possible to discover from behind the interface of an unknown body, embedded in an unknown world? In previous work [1] we presented an algorithm that can deduce the dimensionality of the outside space in which it is embedded, by making random movements and studying the intrinsic properties of the relation linking outgoing motor orders to

resulting changes of sensory inputs (the so called sensorimotor law [2]).

In the present article we provide a more advanced mathematical overview together with a more robust algorithm, and we also present a multimodal simulation.

The mathematical section provides a rigorous treatment, relying on concepts from differential geometry, of what are essentially two very simple ideas. The first idea is that transformations of the organism-environment system which leave the sensory inputs unchanged will do this independently of the code or the structure of sensors, and are in fact the only aspects of the sensorimotor law that are independent of the code (property 1). In a single given sensorimotor configuration the effects of such transformations induce what is called a tangent space over which linear algebra can be used to extract a small number of independent basic elements, which we call “measuring rod”. The second idea is that there is a way of applying these measuring rods globally (property 2) so as to discover an overall substructure in the set of transformations that the organism-environment system can suffer, and that leave sensory inputs unchanged. Taken together these ideas make it possible, if the sensory devices are sufficiently informative, to extract an algebraic group structure corresponding to the intrinsic properties of the space in which the organism is embedded.

The simulation section is for the moment limited to an implementation of the first idea. It presents briefly the main steps of an implementation giving access to the measuring rods, and presents the results of its application to a virtual rat with mixed visual, auditory and tactile sensors (see Figure 2). The group discovered reveals the properties of the Euclidian space implicit in the equations describing the physics of the simulated world.

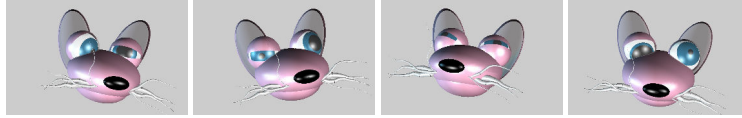


Figure 1: The virtual organism used for the simulations. Random motor commands produce random changes in the rat’s body configuration, involving uncoordinated movements of the head, changes in the gaze direction, and changes in the aperture of the eyelids and diaphragms.

2 Mathematical formulation

Let us note S the sensory inputs, and M the motor outputs. They are the only things the algorithm can access. Let us note P the configurations of the body controlled by the algorithm and E the configurations of the environment.

We will assume that the body position is controlled by the multidimensional motor outputs through some law φ_a and that the sensory devices together deliver a multidimensional input that is a function φ_b of the configuration of the body and the configuration of the environment:

$$P = \varphi_a(M) \quad \text{and} \quad S = \varphi_b(P, E)$$

We shall write $\varphi(M, E) \stackrel{\text{def}}{=} \varphi_b(\varphi_a(M), E)$, note $\mathcal{S}, \mathcal{M}, \mathcal{P}, \mathcal{E}$ the sets of all S, M, P, E , and assume that \mathcal{M} and \mathcal{E} are manifolds.

2.1 Isotropy group of the sensorimotor law

Through time, the algorithm will be able to experiment a set of *sensorimotor laws* linking its inputs to its outputs:

$$\varphi(\cdot, \mathcal{E}) \stackrel{\text{def}}{=} \{M \mapsto \varphi(M, E), E \in \mathcal{E}\}$$

These are a set of functions linking S to M , parametrized by the environmental state E . Our goal is to extract from this set something that does not depend on the way the sensory information is provided. In other words something that would be the same for all $h \circ \varphi(\cdot, \mathcal{E})$, where h is an invertible function corresponding to a change of encoding, including changes of the sensory devices (as long as they provide access to the same information).

If we note $Sym(X) \stackrel{\text{def}}{=} \{f : X \rightarrow X, f \text{ one to one mapping}\}$, and consider the isotropy group of φ in $Sym(\mathcal{M} \times \mathcal{E})$ for the composition action :

$$\Gamma(\varphi) = \{f \in Sym(\mathcal{M} \times \mathcal{E}) \text{ such that } \varphi \circ f = \varphi\}$$

then

Property 1 $\Gamma(\varphi_1) = \Gamma(\varphi_2) \Leftrightarrow \exists f \in Sym(\mathcal{S}) \text{ such that } \varphi_1 = f \circ \varphi_2$

Thus $\Gamma(\varphi)$ is invariant by change of encoding, and retains from φ all that is independent of the encoding. This result is easily understood using an example from physics: think of a light sensor with unknown characteristics in a world consisting of a single point light source. The *values* of the measures are very dependent on the sensor, but the fact that they are equal on concentric spheres is an intrinsic property of the physics of the situation ($\Gamma(\varphi)$, in this case, would be the group of rotations) and is independent of the code and of the sensor's characteristics.

But how to understand the transformations f which, first, involve a manifold \mathcal{E} the algorithm does not know, and second that are *invisible* since $\varphi \circ f = \varphi$. We will show that, under one reasonable assumption, there is an algorithm that can discover the Lie algebra of the Lie subgroups of $\Gamma(\varphi)$ that have independent actions over \mathcal{M} and \mathcal{E} , i.e. Lie groups G such that $g(M, E) = (g_1(M), g_2(E))$ for any $g \in G$, with

$$\varphi(g_1(M), g_2(E)) = \varphi(M, E) \quad \forall g \in G \quad (1)$$

2.2 Fundamental vector fields over the sensory inputs

We will assume that the sensory inputs provide enough information to observe univocally the changes of the environment when the exteroceptive sensors do not move. In mathematical form, we will assume that:

Condition 1 *There exists $\mathcal{U} \times \mathcal{V} \subset \mathcal{M} \times \mathcal{E}$ such that $\varphi(M, \cdot)$ is an injective immersion from \mathcal{V} to \mathcal{S} for any $M \in \mathcal{U}$*

Under this condition, the immersion theorem implies that $\varphi(M, \mathcal{V})$ is a manifold for any $P \in \mathcal{U}$ and that $\varphi(M, \cdot)$ is a diffeomorphism from \mathcal{V} to $\varphi(M, \mathcal{V})$. We shall write $\varphi^{-1}(M, \cdot)$ its inverse. Choosing $M_0 \in \mathcal{U}$, it is thus possible to define an action ϕ^{M_0} of G over the manifold $\varphi(M_0, \mathcal{V})$:

$$\phi^{M_0}(g, S) \stackrel{\text{def}}{=} \varphi(M_0, g_2(\varphi^{-1}(M_0, S))) \quad \forall S \in \varphi(M_0, \mathcal{V})$$

As a consequence (see for instance [3]), for any left invariant vector field X on G there is an associated fundamental vector field X^S on $\varphi(M_0, \mathcal{V})$ ¹ :

$$X^S(S) \stackrel{\text{def}}{=} \frac{d}{dt} \phi^{M_0}(e^{-tX}, S)|_{t=0} \quad \forall S \in \varphi(M_0, \mathcal{V})$$

¹Please note that to avoid heavy notations we have written X^S instead of $X^{\varphi(M_0, \mathcal{V})}$.

The key point for us is that this whole vector field can be discovered experimentally by the algorithm from one vector alone : let us suppose the algorithm knows the one vector $\frac{d}{dt} \phi_1(e^{-tX}, M_0)|_{t=0} \in T\mathcal{M}|_{M_0}$ (the tangent space of \mathcal{M} at M_0), that we will call a measuring rod. Then it can construct a motor command $M_X(t)$ such that

$$M_X(0) = M_0 \quad \text{and} \quad \dot{M}_X(0) = -\frac{d}{dt} \phi_1(e^{-tX}, M_0)|_{t=0}$$

and observe the fundamental field, thanks to the property:

Property 2 $X^S(S) = \frac{d}{dt} \varphi(M_X(t), \varphi^{-1}(M_0, S))|_{t=0} \quad \forall S \in \varphi(M_0, \mathcal{V})$

Indeed the movements of the environment reveal a sub-manifold $\varphi(M_0, \mathcal{V})$ of the manifold \mathcal{S} of all sensory inputs, and this means they allow to transport the sensory image of the given measuring rod over this sub-manifold : $X(S)$ is the time derivative of the sensory inputs at $t = 0$ in the movement implied by the motor command M_X in that configuration of the environment yielding S at $t = 0$.

The fundamental vector fields are the key to our problem because [3] :

$$[X^S, Y^S] = [X, Y]^S$$

where the left term uses the bracket of the vectors fields on $\varphi(M_0, \mathcal{V})$ and the right term uses the bracket in the Lie algebra of G . Thus clearly we can get insight into the properties of the latter by the study of these fields. If the action ϕ^{M_0} is effective (and it is possible to show that for any G there is a subgroup such that it is), we have the additional properties:

1. $X \mapsto X^S$ is an injective Lie algebra morphism: we can understand the whole Lie algebra of G through the Lie bracket over the fundamental vector fields
2. G is diffeomorphic to the group of finite compositions of fundamental flows : any element g of G can be written as $g = e^{X_1} e^{X_2} \dots e^{X_k}$, and

$$\phi^{M_0}(g, S) = \phi^{M_0}(e^{X_1}, \phi^{M_0}(e^{X_2}, \dots \phi^{M_0}(e^{X_k}, S)))$$

2.3 Discovery of the measuring rods

Thus the question is: how can the algorithm come to know the measuring rods? If $\frac{\partial \varphi_b}{\partial P}(\varphi_a(M_0), E_0)$ is for example injective (meaning that there are no small movements from the position P_0 leaving the sensory inputs unchanged), then it can be demonstrated that:

Property 3 $\frac{\partial \varphi}{\partial M}(M_0, E_0) [\dot{M} - \dot{M}_X] = 0 \Rightarrow \frac{d}{dt} \varphi(M(t), \cdot)|_{t=0} = X^S(\varphi(M_0, \cdot))$

This means that the particular choice of one vector of $T\mathcal{M}|_{M_0}$ among those that have the same sensory image as a given measuring rod is of no importance for the construction of the associated vector field. Consequently, the search for the measuring rods becomes the search for their sensory image, which form a linear subspace of the intersection of the tangent spaces of $\varphi(M_0, \mathcal{V})$ and $\varphi(\mathcal{U}, E_0)$ (as a direct consequence of property 2):

$$\forall X \quad \frac{\partial \varphi}{\partial M}(M_0, E_0) \frac{d}{dt} \phi_1(e^{-tX}, M_0)|_{t=0} \in T\varphi(M_0, \mathcal{V})|_{S_0} \cap T\varphi(\mathcal{U}, E_0)|_{S_0}$$

But what about the rest of the intersection? Reciprocally, it can be shown that:

Property 4 *Any measuring rod that has a sensory image in the intersection of the tangent spaces of $\varphi(M_0, \mathcal{V})$ and $\varphi(\mathcal{U}, E)$ for any $E \in \mathcal{V}$ reveals a monodimensional subgroup of transformations over \mathcal{V} that is invariant under any change of encoding.*

3 Simulation

3.1 Description of the virtual rat

We have applied these ideas to a virtual body satisfying the different necessary conditions for the theory to be applied. Though our approach would also apply to the situation where the sensorimotor law involves time-varying functions, for simplicity here we shall take the restricted case where S and M are linked by an instantaneous relationship. We thus implemented a rat's head with instantaneous reactions so that $M \in \mathcal{R}^m$ and $S \in \mathcal{R}^s$. In the simulation, m and s have been arbitrarily assigned the value 300.

The head had visual, auditory and tactile input devices (see Figure 2). The visual device consisted of two eyes, each one being constituted by 40 photosensitive cells randomly distributed on a planar retina, one lens, one diaphragm (or pupil) and two eyelids. The images of the 9 light sources constituting the environment were projected through the lens on the retina to locally stimulate photosensitive cells, with a total influx related to the aperture of the diaphragm and the eyelids. The auditory device was constituted by one amplitude sensor in each of the two ears, with a sensitivity profile favoring auditory sources with azimuth and elevation 0° with respect to the orientation of the head. The tactile device was constituted by 4 whiskers on each side of the rat's jaw, that stuck to an object when touching it, and delivered a signal related to the shift from rest position. The global sensory inputs of dimension 90 (2×40 photosensors plus 2 auditory sensors plus 8 tactile sensors) were delivered to the algorithm through a linear mixing of all the signals delivered by these sensors, using a random matrix $W_S \in \mathcal{M}(s, 90)$ representing some sensory neural encoding in dimension $s = 300$.

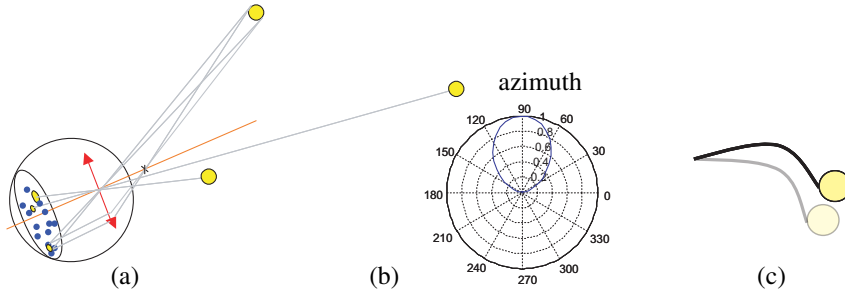


Figure 2: The sensory system. (a) the sensory part of both eyes is constituted of randomly distributed photosensitive cells (small dark dots). (b) the auditory sensors have a gain profile favoring sounds coming from the front of the ears. (c) tactile devices stick to the sources they come into contact with.

The motor device was as follows. Sixteen control parameters were constructed from linear combinations of the motor outputs of dimension $m = 300$ using a random matrix $W_M \in \mathcal{M}(16, m)$ representing some motor neural code. The configuration of the rat's head was then computed from these sixteen variables in this way: six parameters controlled the position and orientation of the head, and, for each eye, three controlled the eye orientation plus two the aperture of the diaphragm and the eyelids. The whiskers were not controllable, but were fixed to the head.

In the simulation we used linear encoding W_S and W_M in order to show that the algorithm worked even when the dimension of the sensory and motor vectors was high. Note first however that any, even non-linear, continuous high-dimensional function could have been used instead of the linear mixing matrices. More important, note that even when linear mixing is used, the sensorimotor law is highly *nonlinear*: the sensors deliver signals that

are not linear with respect to the configuration of the rat's head, and this configuration is itself not linear with respect to the motor outputs.

3.2 The algorithm

The first important result of the mathematical section was that the sensory images of the measuring rods are in the intersection between the tangent space of the sensory inputs observed when issuing different motor outputs while the environment is immobile, and the tangent space of the sensory inputs observed when the command being issued is constant.

In the present simulation we will only be making use of this point, but keep in mind that the second important result was the relation between the fundamental vector fields and these measuring rods. This implies that the tangent vectors we are going to find by an experiment for a given sensory input $S_0 = \varphi(M_0, E_0)$ can be transported in a particular way over the whole sub-manifold $\varphi(M_0, \mathcal{V})$, thereby generating the sensory consequences of any transformation of \mathcal{E} associated with the Lie subgroup of $\Gamma(\varphi)$ whose measuring rods have been found.

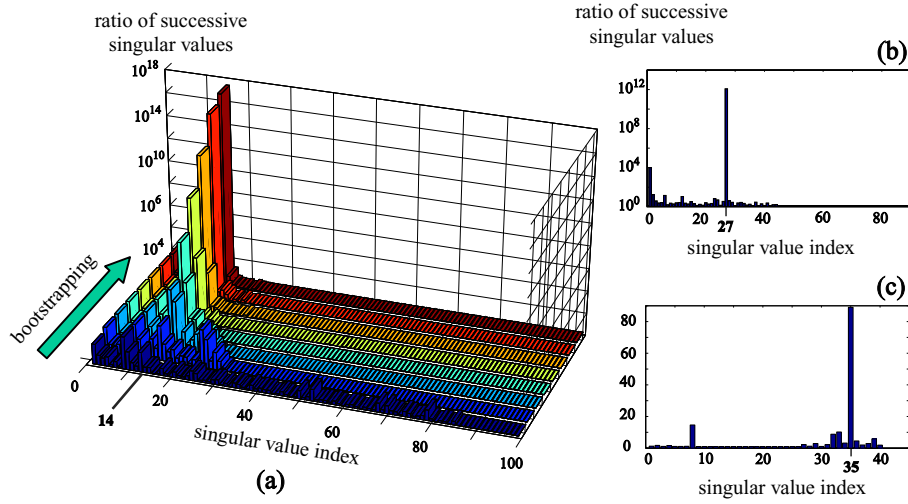


Figure 3: Amplitudes of the ratio of successive singular values of : (a) the estimated tangent sensorimotor law (when E is fixed at E_0) during the bootstrapping process; (b) the matrix corresponding to an estimated generating family for the tangent space to the manifold of sensory inputs observed when M is fixed at M_0 ; (c) the matrix constituted by concatenating the vectors found in the two previous cases. The nullspaces of the two first matrices reflect redundant variables; the nullspace of the last one is related to the intersection of the two first tangent spaces (see equation 2). The graphs show there are 14 control parameters with respect to the body, and 27 variables to parametrize the environment (see text). The nullspace of the last matrix leads to the computation of an intersection of dimension 6 reflecting the Lie group of Euclidian transformations $SE(3)$ (see text).

In [1], the simulation aimed to demonstrate that the dimensions of the different vector spaces involved were accessible. We now present a simulation that goes beyond this by estimating these vector space themselves, in particular $T\varphi(M_0, \mathcal{V})|_{S_0} \cap T\varphi(\mathcal{U}, E_0)|_{S_0}$, in the case of multimodal sensory inputs and with a robust algorithm. The method previously used to estimate the first tangent space, and more specifically its dimension, indeed required

an unrealistic level of accuracy. One of the reasons was the poor behavior of the Singular Value Decomposition when dealing with badly conditioned matrices. We have developed a much more stable method, that furthermore uses time derivatives as a more plausible way to estimate the differential than multivariate linear approximation. Indeed, the nonlinear functional relationship between the motor output and the sensory inputs implies an exact linear relationship between their respective time derivative at a given motor output M_0

$$S(t) = \varphi(M(t), E_0) \Rightarrow \dot{S}(0) = \frac{\partial \varphi}{\partial M}(M_0, E_0) \dot{M}(0)$$

and this linear relationship can be estimated as the linear mapping associating the $\dot{M}(0)$, for any curve in the motor command space such that $M(0) = M_0$, to the resulting $\dot{S}(0)$. The idea is then to use bootstrapping to estimate the time derivative of the “good” sensory input combinations along the “good” movements so that this linear relation is diagonal and the decomposition unnecessary : the purpose of the SVD used at each step is to provide an indication of what vectors seem to be of interest. At the end of the process, when the linear relationship is judged to be sufficiently diagonal, the singular values are taken as the diagonal elements, and are thus estimated with the precision of the time derivative estimator. Figure 3a presents the evolution of the estimated dimension of the tangent space during this bootstrapping process.

Using this method in the first stage of the experiment when the environment is immobile makes it possible for the algorithm, at the same time as it finds a basis for the tangent space, to calibrate the signals coming from the head : it extracts sensory input combinations that are meaningful as regards its own mobility. Then during a second stage, using these combinations, it estimates the tangent space to sensory inputs resulting from movement of the environment while it keeps its motor output fixed at M_0 . Finally, using the tangent spaces estimated in these two stages, it computes their intersection : if TS_M is a matrix containing the basis of the first tangent space, and TS_E a basis of the second tangent space, then the nullspace of $[TS_M, TS_E]$ allows to generate the intersection of the two spaces:

$$[TS_M, TS_E]\lambda = 0 \Rightarrow TS_M \lambda_M = -TS_E \lambda_E \quad \text{where } \lambda = (\lambda_M^T, \lambda_E^T)^T \quad (2)$$

To conclude, using the pseudo-inverse of the tangent sensorimotor law, the algorithm computes measuring rods that have a sensory image in that intersection; and this computation is simple since the adaptation process made the tangent law diagonal.

3.3 Results

The Matlab code of the simulation can be downloaded at <http://nivea.psychol.univ-paris5.fr/~philipona> for further examination of the details of the algorithm and the results.

Figure 3a demonstrates the evolution of the estimation of the ratio between successive singular values. The maximum of this ratio can be taken as the frontier between significantly non-zero values and zero ones, and thus reveals the dimension of the tangent space to the sensory inputs observed in an immobile environment. There are indeed 14 effective parameters of control of the body with respect to the sensory inputs: from the 16 parameters described in section 3.1, for each eye the two parameters controlling the aperture of the diaphragm and the eyelids combine in a single effective one characterizing the total incoming light influx.

After this adaptation process the tangent space to sensory inputs observed for a fixed motor output M_0 can be estimated without bootstrapping as shown, as regards its dimension ($27 = 9 \times 3$ for the 9 light sources moving in a three dimensional space), in Figure 3b. The intersection is computed from the nullspace of the matrix constituted by concatenation

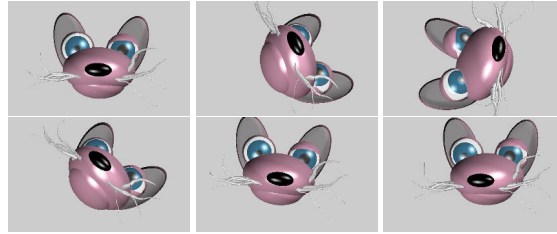


Figure 4: The effects of motor commands corresponding to a generating family of 6 independent measuring rods computed by the algorithm. They reveal the control of the head in a rigid fashion. Without the Lie bracket to understand commutativity, these movements involve arbitrary compositions of translations and rotations.

of generating vectors of the two previous spaces, using equation 2. This nullspace is of dimension $41 - 35 = 6$, as shown in Figure 3c. Note that the graph shows the *ratio* of successive singular values, and thus has one less value than the number of vectors. Figure 4 demonstrates the movements of the rat's head associated with the measuring rods found using the pseudoinverse of the sensorimotor law. Contrast these with the non-rigid movements of the rat's head associated with random motor commands of Figure 1.

4 Conclusion

We have shown that sensorimotor laws possess intrinsic properties related to the structure of the physical world in which an organism's body is embedded. These properties have an overall group structure, for which smoothly parametrizable subgroups that act separately on the body and on the environment can be discovered. We have briefly presented a simulation demonstrating the way to access the measuring rods of these subgroups.

We are currently conducting our first successful experiments on the estimation of the Lie bracket. This will allow the groups whose measuring rods have been found to be decomposed. It will then be possible for the algorithm to distinguish for instance between translations and rotations, and between rotations around different centers.

The question now is to determine what can be done with these first results: is this intrinsic understanding of space enough to discover the subgroups of $\Gamma(\varphi)$ that do not act both on the body and the environment: for example those acting on the body alone should provide a decomposition of the body with respect to its articulations.

The ultimate goal is of course to show that there is a way of extracting *objects* in the environment from the sensorimotor law, even though nothing is known about the sensors and effectors.

References

- [1] D. Philipona, K. O'Regan, and J.-P. Nadal. Is there something out there ? Inferring space from sensorimotor dependencies. *Neural Computation*, 15(9), 2003.
- [2] J. K. O'Regan and A. Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):883–917, 2001.
- [3] T. Masson. *Géométrie différentielle, groupes et algèbres de Lie, fibrés et connexions*. LPT, Décembre 2001. URL : http://qcd.th.u-psud.fr/page_perso/Masson/coursgeodiff.pdf.

6.3 Compléments aux articles

Il me semble intéressant d'apporter deux compléments à ces articles, concernant le cadre conceptuel dans lequel ils se placent : un commentaire sur la légitimité d'utiliser le langage de la géométrie différentielle pour modéliser les dépendances sensorimotrices, et un second sur le rapport entre l'approche prise et la problématique fondamentale, dans le domaine de la perception, de l'inférence.

6.3.1 Le choix du langage de la géométrie différentielle

Le parti pris d'un discours sur l'espace sensorimoteur dans le langage de la géométrie différentielle peut sembler contradictoire avec les caractéristiques du système biologique que nous sommes.

Si l'on considère le système nerveux à une certaine échelle, en effet, un aspect essentiel de la dynamique de l'état des neurones réside dans les temps inter-spike. Or, s'il existe des situations où ces temps sont reproductibles avec une précision inférieure à la milliseconde, notamment dans le cas de stimuli transients ([Mainen and Sejnowski, 1995](#); [Gerstner, 2000](#); [Fellous et al., 2001](#)), ils démontrent plus couramment une grande variabilité, sans doute en raison d'un manque de fiabilité des transmission synaptiques ([Calvin and Stevens, 1967](#); [Korn and Faber, 1990](#)), ou de dynamiques neuronales chaotiques inhérentes aux réseaux exciteur/inhibiteur ([van Vreeswijk and Sompolinsky, 1996](#); [Korn and Faure, 2003](#)). Cette absence de déterminisme à un certain niveau de notre organisme peut paraître contradictoire avec l'utilisation de la géométrie différentielle, et plutôt plaider en faveur d'une approche fondamentalement probabiliste.

La variabilité des signaux neuronaux est un aspect important de notre fonctionnement. Elle semble par exemple expliquer les trajectoires stéréotypées des mouvements des yeux lors des saccades, ou de la main lors des tâches de pointage : ces trajectoires seraient celles minimisant l'incertitude sur la position finale ([Harris and Wolpert, 1998, 2006](#)). Ou encore, du côté sensoriel : les neurones du cortex sensoriel primaire représentant de manière probabiliste les stimuli, la plus ou moins grande fiabilité des informations présentes dans les populations de neurones permettrait d'expliquer la pondération de ces informations observée dans certains jugements perceptif ([Deneve et al., 1999](#)).

Toutefois cette variabilité ne semble pas incompatible avec la tentative de com-

prendre les dépendances sensorimotrices dans le cadre de la géométrie différentielle : ainsi, la thermodynamique exprime dans ce langage des contraintes significatives existant sur le fonctionnement de systèmes de nature essentiellement probabiliste. Cela pose par contre le problème de savoir quelles sont les variables représentées dans le modèle, des lois sensorimotrices régulières et déterministes n'existant sans doute pas au niveau des variables physiologiques associées à des neurones individuels.

Pour approfondir l'analogie de la thermodynamique, l'équation $PV = nRT$ des gaz parfaits peut être vue comme une dépendance température-pression indexée par le volume, et il s'agit d'une relation différentiable et déterministe entre des grandeurs qui ne sont attachées à aucune molécule individuelle, seulement à un ensemble de molécules. De même, les variables sensorielles et motrices entre lesquelles existent des dépendances (l'équivalent de T et P), et les variables indiquant les dépendances sensorimotrices dans lesquelles il est possible de s'engager (l'équivalent de V), et donc permettant l'adaptation adéquate de l'action et du raisonnement, n'ont pas *a priori* à être des variables physiologiques du type potentiel de membrane ou temps inter-spike pour des neurones individuels : cela peut être, entre autre, des variables macroscopiques attachées à différents systèmes physiologiques. On pourrait même envisager qu'il s'agisse de variables de différentes nature, comme le rythme de décharge moyen et la variabilité de ce rythme sur une population : l'important est que ces grandeurs contraignant le fonctionnement de notre organisme soient connectées les unes aux autres par une relation du type des gaz parfaits.

Bien sûr, toutes les grandeurs physiologiques, même le potentiel de membrane ou les temps inter-spike, sont en fait déjà elles-mêmes des grandeurs macroscopiques au regard des processus biochimiques sous-jacents. Il n'y a pas de raison *a priori* d'attribuer un statut particulier, pour expliquer nos jugements perceptifs, à ces grandeurs là précisément : il ne peut y avoir qu'une raison empirique, et si d'autres grandeurs, régulières celles là, se montrent plus pertinentes pour rendre compte de contraintes globales de fonctionnement de notre organisme, ces grandeurs sont tout aussi fondamentales pour expliquer l'origine de notre discours sur l'expérience.

6.3.2 Approche sensorimotrice et inférence

Un grand nombre de données sur la perception et les illusions trouvent une explication dans un cadre inférentiel, en particulier Bayésien. Ces données concernent par

exemple la perception du relief (Mamassian and Landy, 2001), du mouvement (Weiss et al., 2002), des couleurs (Yang and Maloney, 2001), ou encore les effets de groupement (Geisler et al., 2001). L'explication proposée pour tous ces cas est que nous estimons, étant données des informations sensorielles ambiguës et une certaine probabilité d'occurrence des propriétés de l'environnement, être confrontés à la propriété de l'environnement la plus probable.

Il me semble que l'approche sensorimotrice et l'approche Bayésienne sont naturellement complémentaires l'une de l'autre. D'un côté, il se pose en effet dans l'approche sensorimotrice la question d'estimer, à partir de certaines dépendances sensorimotrices dans laquelle l'organisme peut vérifier pouvoir s'engager, dans quelles autres dépendances sensorimotrices il doit pouvoir s'engager. Cette estimation doit naturellement pouvoir se faire dans un cadre Bayésien. Et de l'autre côté, considérée isolément, l'approche Bayésienne souffre d'un réalisme naïf : elle pose le problème de la perception comme celui de percevoir les propriétés de l'environnement, en passant sous silence qu'aucun principe n'est donné pour déterminer ce qui doit compter comme étant les propriétés de l'environnement perçues.

Cette dernière difficulté, et ses conséquences, est illustrée par une théorie Bayésienne de la constance des couleurs (Brainard and Freeman, 1997; Brainard et al., 2006), qui suggère que le système nerveux estimerait *le spectre* de l'illuminant, lorsqu'il n'est d'aucune manière évident qu'un tel objet, utile pour rendre compte d'expériences impliquant un appareillage physique particulier, le soit également pour rendre compte du fonctionnement de notre propre organisme dans son environnement. En fait, il est par exemple absurde d'imaginer que le spectre en dehors du domaine de sensibilité des photorécepteurs affecte notre fonctionnement, en tout cas d'une manière similaire au spectre dans le domaine visible. Par analogie, en considérant que nous ne disposons que de quelque types de photorécepteurs rétinien, dont le fonctionnement ne peut donc distinguer toute une variété de compositions spectrales différentes, l'idée d'une estimation du spectre détaillé semble douteuse. Si une approche Bayésienne de la constance des couleurs paraît très naturelle, le coeur de la question ne me semble pas être le détail de l'algorithme d'estimation, mais la proposition *de ce qui est estimé*.

Un autre exemple est que, comme on l'a déjà dit en introduction de ce chapitre, l'idée que nous vivons dans un monde tridimensionnel euclidien habité par des objets ayant certaines propriétés, pour cohérente qu'elle soit, dépend au minimum de la

donnée d'une échelle. Pourquoi devrait-il être évident qu'il faut appliquer l'approche Bayésienne à des variables physiques telles que l'orientation, la taille, le poids, la vitesse, lorsque le "vrai" monde, celui qui nous est décrit par la physique fondamentale, est un monde de particules étranges où ces propriétés n'existent pas ? Quels sont les principes qui doivent permettre de déterminer ce que sont les propriétés macroscopiques estimées ?

L'approche Bayésienne, semble-t-il, n'a pas d'autre choix que d'appeler propriétés de l'environnement ces propriétés qui ont une pertinence au niveau du fonctionnement global du système que nous sommes dans notre environnement. Et cela est largement compatible avec l'approche sensorimotrices, qui vise, il me semble, à proposer une spécification de cette notion de propriétés pertinentes.

6.4 Critique des articles

Il me paraît également intéressant de donner sur les deux articles précédents des commentaires un peu plus critiques, vis-à-vis des interprétations auxquelles ils laissent place, et vis-à-vis de certaines limites de l'argumentation.

6.4.1 Sur l'interprétation de l'argument

Le premier point qui appelle un commentaire est que ces articles peuvent donner l'impression de vouloir démontrer quelque chose au moyen d'une théorie mathématique élaborée. Or il n'est pas inutile de préciser explicitement qu'il n'y a aucune forme de démonstration (en tout cas de démonstration non triviale) : la formalisation mathématique n'est utilisée dans ces articles que comme moyen d'expression, pour *décrire* dans un langage supposé non ambigu ce qui pourrait être entendu par les termes "dépendances sensorimotrices", "propriétés des dépendances", "structure", "non équivalence", etc.

Le second point est que les deux articles visaient à montrer par la pratique, au travers de simulations, que les propriétés mises en avant dans un discours plus théorique sont effectivement des propriétés existant dans des dépendances sensorimotrices possédant une certaine forme de réalisme. L'évocation d'une méthode permettant de révéler ces propriétés peut, toutefois, donner lieu à deux malentendus : l'idée que l'intérêt des articles est dans la description d'un algorithme précis d'identification en tant que

telles de certaines caractéristiques des dépendances, et l'idée qu'il était proposé que la sensation de spatialité découle d'un tel savoir, "acquis par le cerveau".

Or, en tant que méthode, ce qui est décrit dans ces articles est simpliste, souffre de différents problèmes de fiabilité et de stabilité : l'objectif réel était de montrer qu'il est cohérent de dire que la tridimensionalité de l'espace, par exemple, est une propriété des dépendances, pas de donner un algorithme permettant d'estimer cette dimension. En effet, quel intérêt pour un organisme d'estimer en tant que telle cette dimension ? L'argument est que la sensation de spatialité est la capacité pratique *d'exploiter* certaines analogies sensorimotrices entre différentes situations perceptives (un savoir faire), analogies contraintes par un fait tel que la tridimensionalité, pas d'avoir une forme de "savoir sur" ces contraintes. Plus finement : ce n'est pas un "savoir sur" de l'individu, comme en atteste le fait que la tridimensionalité de l'espace est un concept vague pour un non-mathématicien, mais ce n'est pas non plus un "savoir sur du cerveau" : la capacité d'exploiter certaines analogies sensorimotrices est une capacité rendue possible par le fonctionnement du cerveau (et du reste de notre corps), mais que l'on doit dire, selon l'usage du langage, être un savoir faire *de l'individu*.

En fait, ce que peut bien apporter la métaphore d'un cerveau personnifié, présente dès (O'Regan and Noë, 2001) avec l'analogie de l'équipe d'ingénieurs dans un sous-marin, une fois que tout ce qu'elle ne peut pas être a été mis de côté, je ne le sais pas. Ce qu'elle ne peut pas être, c'est par exemple l'affirmation qu'il y a un savoir du cerveau sur les dépendances sensorimotrices comme il y a un savoir des ingénieurs sur la manière dont se comporte le sous-marin qu'ils contrôlent, i.e. une capacité de discourir et d'agir en relation avec ce discours. Ou l'affirmation qu'il y a le problème suivant : différents signaux sensoriels que le cerveau doit analyser, par exemple identifier comme étant des signaux de nature visuelle ou auditive.

Sur cette seconde affirmation, en effet, s'il est absurde au sens littéral d'imaginer que le cerveau observe les signaux rétiniens comme un capitaine de sous-marin observe les images d'un périscope, il est également absurde de l'imaginer *au sens figuré*. Cela est absurde parce que la relation du cerveau aux signaux neuronaux, n'ayant rien à voir avec celle d'une personne observant un ensemble d'objets, il y a une *inadéquation structurelle* de la métaphore. Il y a ainsi un sens clair à dire que deux objets sont identiques vis-à-vis d'une personne dès lors que, lorsqu'elle est face à l'un ou à l'autre, elle se trouve, localement, dans la même situation (elle peut s'engager dans un même

ensemble d'interactions sensorimotrices, notamment manipulatoires). A l'inverse, ce que pourrait vouloir dire l'identité ou la non identité de deux signaux neuronaux vis-à-vis du cerveau est pour le moins obscur : le cerveau ne peut pas se tourner pour regarder l'un des signaux, puis l'autre, et se trouver ou pas dans la même situation³. Il n'est pas possible de transposer certains concepts implicitement rattachés au verbe "observer", tels que les concepts d'observer des choses qui peuvent se trouver être identiques, et qui constituent une partie de son sens.

L'histoire de l'équipe d'ingénieur nous interpelle, elle décrit un problème intelligible : les expressions "essayer des commandes motrices", "observer des signaux sensoriels", "déduire", "être naïf", sont des expressions que nous savons utiliser en situation, et que nous savons placer au centre d'autres discours. Seulement, tout le problème est ensuite dans la pertinence de cette histoire comme métaphore du problème de l'expérience lorsqu'il n'est pas possible de donner un sens à l'utilisation de ces expressions, comme on vient de le voir pour l'expression "observer des signaux", avec le sujet grammatical "cerveau". Le cerveau fonctionne, c'est tout. Celui qui sait, celui qui perçoit, c'est l'individu, le référent de ce sujet grammatical qui peut être utilisé dans tel ensemble d'affirmations, dans lequel le terme "cerveau" ne peut pas être utilisé.

6.4.2 Sur l'argument lui-même

En dehors de ces problèmes d'interprétation, les articles précédents souffrent également de nombreuses insuffisances plus "techniques". On peut en citer en particulier deux, pour justifier de la nécessité de proposer une reformulation de l'approche :

1. le terme de compensabilité renvoie à une notion d'adaptabilité d'un ensemble de commandes motrices, et pas seulement d'évolution d'une commande donnée, comme suggéré dans les articles. Il s'agit d'un point conceptuellement important car, me semble-t-il, le coeur de l'approche sensorimotrice est de dire qu'il n'est pertinent de parler d'analogies qu'au niveau des dépendances sensorimotrices (des situations perceptives), pas au niveau des stimulations sensorielles.
2. les propriétés de commutation évoquées dans les articles ne permettent pas de distinguer les changements d'environnement liés exclusivement à la luminance ou exclusivement à un changement d'intensité sonore – un point pourtant essen-

³Un neuroscientifique peut faire cela, et il y a donc un sens à dire que *pour lui* (vis-à-vis de son fonctionnement à lui), les signaux sont identiques ou pas.

tiel puisqu’une telle distinction est nécessaire pour que la théorie propose une explication de l’une des distinctions les plus patentes au niveau de notre expérience : celle de différentes modalités sensorielles.

Considérant une dépendance sensorimotrice φ , modifiée en dépendance φ' après un changements d’environnement, la compensabilité de cette modification relève de l’adaptabilité des variables motrices pour tout un ensemble de valeurs de ces variables, et pas simplement d’un changement (m, m') : s’il était possible pour une certaine valeur m des variables motrices de trouver un ajustement m' telle que $\varphi'(m') = \varphi(m)$, mais que pour toutes les valeurs voisines il n’y ait pas d’ajustement possible, il semblerait tout à fait impropre de dire que la modification de l’environnement peut être annulée. Le terme de compensabilité renvoie bien, intuitivement, à une notion d’adaptabilité d’un ensemble de commandes motrices.

Cela est en phase avec le fait que l’argument de la théorie est bien que lorsqu’un objet se déplace, nous pouvons nous retrouver dans la même situation perceptive modulo une adaptation des commandes motrices : il ne s’agit pas simplement de pouvoir maintenir constant une stimulation sensorielle donnée, mais de dire que toute une structure d’interactions sensorimotrices peut être retrouvée après le déplacement de l’objet.

D’autre part, en n’exploitant que les propriétés de commutativité des changements de dépendance compensables, les deux articles se trouvent incapables de légitimer une distinction entre un changement de luminance seul, et un changement de luminance accompagné d’un changement d’intensité sonore. En effet, ces deux changements ont en réalité exactement le même statut vis-à-vis de la commutation des changements de dépendance : celui de commuter avec tous les autres changements. Or l’approche n’a pas de raison de se focaliser exclusivement sur cette propriété, et il s’agit donc de proposer également d’autres types de propriétés.

6.5 Reformulation de l’approche

Après ces commentaires, ce qui suit présente donc une seconde tentative de formalisation de la notion de structure des dépendances sensorimotrices, corrigeant, je l’espère, certaines des ambiguïtés de présentation et d’argumentation, et qui progresse un peu sur certains points plus essentiels. L’idée générale reste largement identique à

celle des articles, mais propose des évolutions sur les deux difficultés précédemment évoquées.

6.5.1 Univers sensorimoteur

Les activités des neurones moteurs contrôlant la position de la tête, l'orientation des yeux, l'ouverture des iris et la tension des muscle des étriers⁴, déterminent, dans un environnement fixe donné, une unique configuration d'équilibre des muscles, donc en particulier une unique posture de la tête et des yeux, donc la stimulation sensorielle, rétinienne et cochléaire. Dans une telle situation, il existe des variables motrices et sensorielles, à valeur dans des ensembles \mathcal{M} et \mathcal{S} (des ouverts de \mathbb{R}^m et \mathbb{R}^s), telles qu'il existe entre ces variables une dépendance associative $\varphi \in \mathcal{F}(\mathcal{M}, \mathcal{S})$ (Fig. 6.1).

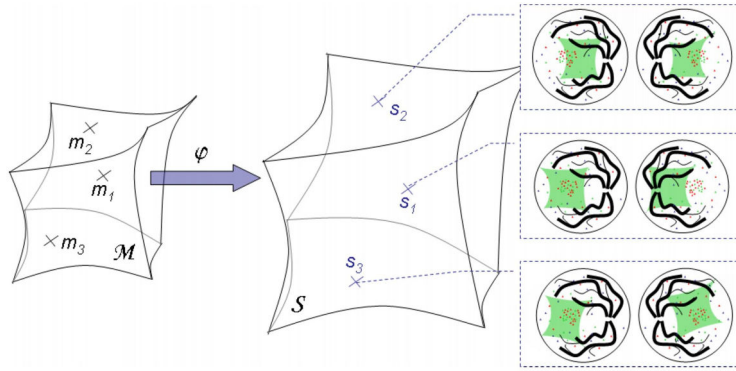


FIG. 6.1 – Dépendance sensorimotrice associative. Chaque activation motrice m_i détermine une stimulation sensorielle s_i . En insert : exemple de stimulations rétinienne produites par un carré de couleur verte, schématisant les obstructions vasculaires, les distortions optiques et la non-uniformité de la distribution des photorécepteurs.

Bien sûr, une description détaillée du système devrait être en réalité plus complexe qu'une simple relation associative : il faudrait par exemple modéliser la manière dont la posture de la tête et des yeux converge temporellement vers l'état d'équilibre déterminé par la commande motrice (les évolutions physiologiques à l'oeuvre dans les fibres musculaires et les inerties de la tête et de yeux), et modéliser la manière dont l'activité sensorielle converge temporellement vers l'état d'équilibre déterminé par la

⁴L'ouverture de l'iris et la tension des muscles des étriers sont des mécanismes élémentaires de protection des cellules sensorielles de la rétine et de l'oreille. Ils sont complétés par des mécanismes neuronaux de modulation de la réponse de ces cellules, capable généralement de jouer sur une plage beaucoup plus importante (l'iris peut modifier d'un facteur x l'intensité lumineuse pénétrant dans l'oeil, alors que différents processus moléculaires modulent d'un facteur y la sensibilité des photorécepteurs.).

posture courante de la tête et des yeux (la cinétique chimique au niveau des photorécepteurs et les interactions neuronales rétinienne, la dynamique mécanique au niveau des cellules ciliées). C'est seulement lorsque les temps caractéristiques des dynamiques motrices et sensorielles sont courts, devant les temps caractéristiques de détermination des commandes motrices, qu'une dépendance associative de la forme précédente décrit l'essentiel du fonctionnement du système⁵.

Il faut noter que, dans le cas des mouvements oculaires, la considération de ce type de dépendance est à la limite d'être légitime : les temps caractéristiques des dynamiques d'intégration rétinienne sont de l'ordre du dixième de seconde, la durée d'une saccade entre le début du mouvement de l'oeil et sa stabilisation peut varier du centième au dixième de seconde (selon la longueur de la saccade), et nous initions en moyenne entre trois et cinq saccades par seconde.

Lorsque l'environnement change, la dépendance sensorimotrice est modifiée. Un ensemble de configurations potentielles pour l'environnement – par exemple, un ensemble de configurations spatiales de quelques objets en face de l'observateur, un ensemble de caractéristiques de l'illuminant et du son ambiant –, déterminent ainsi une *variété* de dépendances (Fig. 6.2) : $\Phi \subset \mathcal{F}(\mathcal{M}, S)$.

Là encore, il s'agit d'une simplification importante. La variété des dépendances dans lesquelles le système sensorimoteur qui a été décrit peut réellement se trouver engagé est bien sûr de beaucoup plus grande dimension que l'espace des possibles considéré dans cet exemple. En fait, si l'on considère le seul système visuel et des environnements constitués par un ensemble de photographies, on voit que l'ensemble des dépendances possibles est aussi riche que l'ensemble de photographies. Dans la mesure où l'on peut même considérer un ensemble de scènes tridimensionnelles formant un ensemble encore plus grand que celui des photographies, il apparaît même que la diversité des dépendances dans lesquelles le système visuel peut se trouver engagé est plus grand que l'ensemble des images.

Considérer une variété d'environnements aussi restreinte que celle juste évoqué manque donc profondément de réalisme, et il faut prendre cet exemple comme un sujet d'inspiration pour des développements ultérieurs, pas comme une théorie très raffinée.

⁵Dans le cas contraire, il se pourrait en effet que les changements de commande motrice s'enchaînent à un rythme tel que la posture n'ait jamais le temps de converger vers l'équilibre déterminé par l'activité neuronale motrice, et que l'activité neuronale sensorielle n'ait jamais le temps de converger vers l'équilibre déterminé (étant donné l'environnement) par la posture courante. Il semblerait dans ce cas douteux de tirer des conclusions de la relation pouvant exister entre les équilibres moteurs et sensoriels.

Il faut noter en particulier qu'il s'agit d'un exemple qui ne retient pas certains problèmes pourtant classiques en vision, tel que l'ambiguïté perceptive entre la distance d'un objet et sa taille (l'existence de cette ambiguïté nécessitant la considération d'un environnement dans lequel se trouve des familles d'objets homothétiques).

On appellera dans la suite *univers sensorimoteur* un triplet $(\mathcal{M}, \mathcal{S}, \Phi)$, et on parlera indifféremment pour les éléments de Φ de dépendance sensorimotrice, d'environnement, ou de situation perceptive : du point de vue proposé, une configuration de l'environnement ou une situation perceptive *est* une dépendance sensorimotrice.

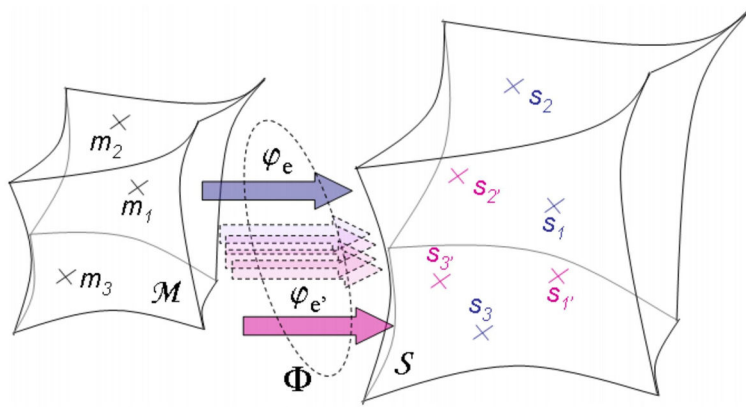


FIG. 6.2 – Univers sensorimoteur. Dans la première configuration de l'environnement, les activations motrices m_i déterminent les stimulations $s_i = \varphi(m_i)$. Dans la seconde configuration de l'environnement, ces mêmes activations motrices déterminent les stimulations $s_{i'} = \varphi'(m_i)$. Φ est l'ensemble des dépendances sensorimotrices résultant d'un ensemble de configurations de l'environnement.

L'objectif de ce qui suit est de montrer que, dans un univers sensorimoteur, toutes les dépendances et tous les changements de dépendance ne sont pas équivalents les uns aux autres : certaines dépendances, certains changements de dépendance, ont des propriétés (de compensabilité, d'invariance, de correspondances,...) catégoriquement différentes les uns des autres, et l'argument est que cela justifie une catégorisation des situations perceptives et de leurs changements.

6.5.2 Changements de dépendance compensables.

L'idée de la compensabilité est que, dans un univers sensorimoteur, certains changements de dépendance sont plus "superficiels" que d'autres : les dépendances sensorimotrices restent identiques à une adaptation près des variables motrices, ce qui n'est

pas le cas des autres. Dans l'exemple considéré, le déplacement rigide d'un objet est un changement de dépendance qui peut être compensé par une adaptation adéquate des variables motrices, alors que le changement de dépendance résultant d'une modification de sa forme ne le peut pas (Fig. 6.3).

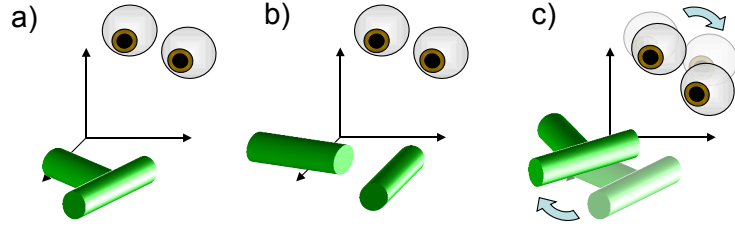


FIG. 6.3 – Changements de dépendance fondamentalement distincts. a) situation perceptive initiale, b) changement de dépendance non compensable, c) changement de dépendance compensable : il suffit d'adapter de manière adéquate les commandes motrices pour retrouver la dépendance initiale

Il est clair, cependant, qu'il ne faut pas espérer pouvoir adapter toutes les commandes motrices : il n'est ainsi pas possible d'adapter les commandes correspondant à une torsion maximum de la tête pour compenser le mouvement d'une cible dans cette direction. Il faut par conséquent nécessairement considérer l'adaptabilité des variables motrices sur *une partie* $U \subset \mathcal{M}$ de leur domaine de définition.

Pour résumer, on dira formellement qu'un changement de dépendance (φ, φ') dans un univers sensorimoteur est *compensable*⁶ par l'adaptation motrice h s'il existe une application $h \in \mathcal{F}(U, \mathcal{M})$ telle que $\varphi' \circ h = \varphi|_U$.

Si les variables sensorielles considérées sont, comme dans l'exemple considéré, l'ensemble des potentiels de membrane des cellules photoréceptrices et des cellules ciliées internes, il faut noter que la compensabilité du déplacement d'un objet dépend du fait que cet objet se trouve devant un fond uniforme, avec un éclairage non directionnel, et qu'il n'y a pas de source sonore localisée.

C'est cette situation, particulièrement simple et intuitive, mais également particulièrement artificielle, qui sert de base plusieurs fois dans la suite pour préciser la notion de compensabilité. Cela ne doit pas suggérer que l'approche est limitée à ce genre de contexte, ni que ces variables sensorielles là ont un statut particulier dans la théorie (c.f. section 6.3.1 sur la nature des variables sensorielles et motrices).

⁶En réalité, on devrait préciser U -compensable, mais on raisonnera toujours par rapport au même U .

Il est certain que, lorsqu'un objet bouge dans notre champ visuel, il n'est généralement pas possible, du fait de la présence d'un fond non uniforme et d'éclairages directionnels, de compenser le changement de dépendance par une simple adaptation motrice des positions de la tête – il faudrait également considérer une adaptation sensorielle, ou considérer des variables sensorielles moins informatives que l'ensemble des potentiels de membrane. Il y a là un point essentiel à développer dans le détail, mais le fait est que lorsqu'un objet bouge devant nous, certaines adaptations motrices sont fondamentales pour nous permettre de réaccéder à certaines caractéristiques de l'objet (les propriétés d'une face, etc). Et c'est ce fait qu'il y a véritablement à retenir dans l'exemple proposé, au delà de l'aspect simpliste de ce dernier suggérant qu'il suffit d'une adaptation motrice pour retrouver exactement les mêmes stimulations sensorielles.

6.5.3 Adaptations motrices génériquement compensatoires

Il se peut que, dans un univers sensorimoteur donnée, une adaptation motrice h compense un changement particulier de dépendance (φ, φ') , et seulement celui là. Il se peut également qu'une adaptation h compense des changements de dépendance de manière plus générique.

En effet, dans l'exemple considéré, il se trouve que certaines adaptations particulières des positions de la tête, des ouvertures des iris, ou des contractions des muscles de l'oreille moyenne, sont telles qu'il se trouve toujours un changement de l'environnement qu'elles peuvent compenser. On parlera dans ce cas d'*adaptations motrices génériquement compensatoires* (AMGC, pour simplifier le texte). Dans la suite, formellement, on notera $\mathcal{H} \subset \mathcal{F}(U, \mathcal{M})$ la variété de ces adaptations h , qui vérifient donc : pour toute dépendance sensorimotrice $\varphi \in \Phi$, il existe dans l'univers sensorimoteur un changement de dépendance (φ, φ') compensable par h .

Il est important de noter que la propriété de généricité n'est, doublement, pas une propriété d'unicité : d'une part, des modifications différentes d'une dépendance donnée peuvent être compensables par une même adaptation motrice, et d'autre part, différentes adaptations motrices peuvent se trouver compenser un même changement de dépendance.

Ainsi, dans l'exemple considéré, si l'on considère la situation initiale où un certain objet est dans l'axe du regard et un autre se trouve hors de vision si ce n'est par une tor-

sion maximum de la tête, alors, parce que $U \subsetneq \mathcal{M}$, seul le mouvement du premier objet détermine le fait que le changement de dépendance est compensable par une adaptation motrice $h \in \mathcal{F}(U, \mathcal{M})$: le second objet peut évoluer de diverses manières dans le même temps, et donc produire divers changements de la dépendance sensorimotrice pourtant tous compensables par la même adaptation h .

D'autre part, du fait de la possible redondance des variables motrices vis-à-vis du contrôle moteur, il se peut que deux adaptations motrices différentes h_1 et h_2 soient en fait complètement équivalentes pour compenser les changements de dépendance, c'est-à-dire que tout changement de dépendance (φ, φ') est compensable par h_1 si et seulement si il l'est par h_2 .

6.5.4 Dépendances invariantes

Les AMGC distinguent d'emblée quelques situations perceptives particulières : certaines dépendances d'un univers sensorimoteur peuvent avoir la propriété d'être invariantes par une ou plusieurs de ces adaptations – formellement, ces dépendances sont telles que $\varphi \circ h = \varphi|_U$ pour différentes adaptations $h \in \mathcal{H}$, alors qu'habituellement, ces adaptations modifient la dépendance. Il n'est pas possible d'établir une analogie sensorimotrice entre des dépendances invariantes et des dépendances non-invariantes, ce sont des situations perceptives profondément différentes, elles ne permettent pas de s'engager dans la même structure d'interactions sensorimotrice.

Les dépendances sensorimotrices associées à un objet ayant une symétrie de translation, un axe de révolution, ou une symétrie sphérique, présentent de manière évidente de telles invariances. Mais c'est également le cas d'autres dépendances, de manière moins anodine.

Ainsi, dans l'exemple considéré, les différentes situations perceptives correspondant à l'*obscurité* (et différentes valeurs des paramètres spectraux de l'environnement sonore, en rappelant qu'il n'y a pas de source localisée dans l'exemple) sont invariantes par toutes les AMGC des positions de la tête, des yeux, et des ouvertures des iris. De la même manière, les situations perceptives correspondant au *silence* (et différentes configurations des objets) sont invariantes par l'adaptation motrice des tensions des muscles de l'oreille interne : alors que ces derniers permettent habituellement de moduler la stimulation cochléaire, dans le silence une adaptation de la tension de ces muscles n'a plus d'effet.

Les propriétés d'invariance sensorimotrice de l'obscurité et du silence sont à mettre en regard du fait que ces situations perceptives ont un statut (hétéro)phénoménologique particulier : sans doute les sujets affirment-ils que ces situations perceptives sont spéciales, et sans doute existe-t-il des mots dans la plupart des langages pour exprimer cette distinction primaire entre l'obscurité et la lumière, le silence et le bruit. Ces affirmations ont une pertinence indiscutable : les sujets se trouvent dans des situations perceptives effectivement très particulières.

Selon l'approche sensorimotrice, nous affirmons que l'expérience de l'obscurité et du silence sont des expériences sensoriellement spéciales parce que ces mots renvoient à des situations perceptives objectivement spéciales (l'invariance étant un exemple de telle particularité objective), et pas parce qu'il y a quelque chose de spécial "en nous" – mystère insondable "d'états mentaux" échappant irrémédiablement au monde physique, ou mystère insondé d'un mécanisme cortical aux propriétés encore incomprises et peut-être incompréhensibles.

Cette ligne d'explication peut sembler naïve. Mais il s'agit de la comparer aux argumentations adverses et pas juste en elle-même, à peine ébauchée. Il s'agit de savoir laquelle des trois lignes d'argumentation précédentes (sensorimotrice, dualiste, purement neuronale), fondamentalement distinctes, peut sembler la moins invraisemblable, une fois les détails ajustés, pour formuler une explication scientifique du discours que nous tenons sur l'expérience.

Il faut noter, enfin, que cette approche n'explique pas l'expérience de l'obscurité ou de silence par "l'absence" de stimulation rétinienne ou cochléaire, comme cela semble implicite dans certains raisonnements psychophysiques : sans plus de précisions, l'absence de stimulation est un concept qui n'est pas évident, notre dispositif sensoriel étant en fonctionnement permanent dans le monde.

La question est précisément d'expliquer pourquoi un flash est considéré comme une stimulation "positive" et l'obscurité comme une absence de stimulation. L'interprétation suggérée ici est que c'est parce qu'un flash nous met dans une situation perceptive où nous pouvons nous engager dans des adaptations motrices pour expérimenter le monde, alors que dans l'obscurité nos adaptations motrices sont sans effet.

6.5.5 Correspondances compensatoires

Dans l'univers sensorimoteur considéré, l'existence d'adaptations motrices génériquement compensatoires est accompagnée d'un autre fait remarquable. Partant, par exemple, de la situation initiale où un seul objet est présent dans le champ de vision, un déplacement de cet objet puis un changement de luminosité de la scène, qui sont chacun des changements de l'environnement compensables par des AMGC données, résultent en un changement global qui est lui même compensable par une AMGC. Qui plus est, cette AMGC est déterminée par les deux adaptations motrice compensant chacun des deux changements, indépendamment de la situation perceptive initiale, c'est-à-dire de la position initiale de l'objet, de l'objet en question, de la luminosité initiale, de l'environnement sonore, etc.

Plus généralement, il existe dans l'univers sensorimoteur considéré tout un ensemble de *correspondances compensatoires*, pouvant chacune être représentée par une expression de la forme $h_1 \dots h_k \sim g$, signifiant que si les changements successifs d'une séquence de dépendances $(\varphi_1, \dots, \varphi_{k+1})$ sont compensables par les AMGC h_1, \dots, h_k , alors le changement $(\varphi_1, \varphi_{k+1})$ est un changement lui même compensable, par l'AMGC g .

Il faut noter que l'existence d'un ensemble de correspondances compensatoires ne signifie pas qu'il existe une correspondance (une expression) pour chaque séquence d'AMGC. Dans l'univers sensorimoteur pris en exemple, le changement de dépendance total résultant de deux itérations d'un changement compensable par une rotation maximale de la tête n'est ainsi lui-même compensable par aucune adaptation des commandes motrices du système visuo-moteur.

6.5.6 Adaptations motrices nulles, résidus de commutation

Les AMGC ne sont pas équivalentes les uns aux autres, vis-à-vis des correspondances compensatoires d'un univers sensorimoteur. Par exemple : certaines adaptations motrices h sont telles que $hg \sim g$ pour toute adaptation $g \in \mathcal{H}$, ce qui ne sera pas le cas des autres. Le fait que les adaptations motrices puissent se trouver vérifier des propriétés différentes vis-à-vis des correspondances compensatoires peut ainsi être la base d'une distinction entre différentes catégories d'adaptations, et donc différentes catégories de changements de l'environnement.

Parmi les propriétés pouvant distinguer les AMGC, on peut par exemple définir :

- la notion d'*adaptation motrice nulle* : on dira que l'adaptation motrice g est nulle si $gh \sim h$ et $hg \sim h$ pour toute adaptation $h \in \mathcal{H}$,
- la notion de *résidu de commutation*⁷ : on dira que l'adaptation motrice g est un résidu de commutation du couple d'adaptations motrices (h_1, h_2) s'il existe h tel que $h_1h_2 \sim h$ et $h_2h_1g \sim h$ (on notera : $h_1h_2 \sim h_2h_1g$). On dira que deux adaptations motrices h_1 et h_2 commutent lorsque leurs résidus de commutation sont tous nuls.

Dans l'exemple considéré, il se trouve que ces deux notions suffisent à distinguer, parmi les AMGC, les adaptations motrices compensant, d'une part, à la fois les changements d'intensité lumineuse et sonore de l'environnement, et, d'autre part, les déplacements des objets en translation et en rotation.

Dans l'exemple décrit, les adaptations motrices g compensant les changements de l'environnement autres que le déplacement d'un objet, en effet, peuvent être caractérisées par le fait qu'elles commutent avec tout changement compensable. Ensuite, les adaptations motrices compensant exclusivement le déplacement d'un objet sont exactement l'ensemble des résidus de commutation. Enfin, parmi les changements spatiaux, les adaptations motrices h associées à une translation peuvent être caractérisées par exemple par le fait qu'il est possible de trouver deux adaptations g_1 et g_2 qui commutent avec h mais qui ne commutent pas entre elles.

Sur le fond, on peut bien sûr s'interroger sur la pertinence de la seule propriété de commutation pour expliquer l'origine du discours selon lequel "l'expérience d'un changement de luminosité est différente de celui du déplacement d'un objet dans notre champ de vision". D'autres propriétés auraient pu être évoquées : il n'est ainsi pas fait état explicitement de la question de la réversibilité, ni du fait qu'une rotation induit des trajectoires closes, ni de ce qu'un rétrécissement de l'iris modifie la stimulation d'une manière entièrement déterminée par la stimulation initiale, ce qui n'est pas le cas d'un mouvement des yeux. Il n'y a pas, dans ce qui précède, de volonté d'exhaustivité. L'objectif est bien uniquement de montrer que les AMGC ne sont pas équivalentes les unes aux autres vis-à-vis des correspondances compensatoires, et que ce genre de différences peut être la base d'une distinction entre plusieurs catégories de changements

⁷Cette notion est évidemment une version pauvre de commutateur, étroitement liée à la notion de crochet de Lie.

de situation perceptive.

6.5.7 Limites de changements compensables

Dans l'exemple décrit, les propriétés d'invariance et de commutation n'apporte aucun argument sur des points pourtant marquants de notre expérience : la distinction entre changements d'intensité lumineuse et sonore de l'environnement, ou la distinction entre un accroissement de ces intensités ou leur diminution. La commutativité, en effet, permet de distinguer ces adaptations motrices compensant les changements de l'environnement autres que le déplacement d'un objet, mais toutes ces adaptations sont ensuite équivalentes entre elles vis-à-vis de la commutativité.

Outre leur commutativité vis-à-vis des correspondances compensatoires, les changements de dépendance peuvent également se distinguer par exemple par les propriétés de l'itération d'un même changement. Ainsi, dans l'exemple considéré, les itérations $(\varphi_1, \dots, \varphi_n, \dots)$ d'un changement non spatial compensable par une adaptation motrice donnée se trouvent n'avoir que trois types de comportements (Fig. 6.4) : la suite de dépendances converge soit vers le silence *et* l'obscurité, soit vers le silence *ou* l'obscurité, soit ne converge pas vers une dépendance invariante. Cette situation établit les distinctions juste évoquées : le second comportement caractérise en effet les adaptations motrices compensant exclusivement une baisse de luminosité, ou exclusivement une baisse d'intensité sonore.

Il est intéressant de noter que cette approche, développée autour de la définition de mêmes changements de l'environnement comme étant des changements compensables par une même adaptation motrice, se trouve dans l'exemple considéré compatible⁸ avec les données connues sous le nom de "loi de Weber-Fechner". Cela découle du fait qu'un changement donné d'ouverture de l'iris compense la *multiplication* de l'intensité lumineuse par un facteur donné, et de même pour la tension des muscles de l'oreille interne et l'intensité sonore de l'environnement : un même changement d'environnement (d'intensité) est une multiplication de l'intensité par un même facteur, pas l'addition d'un même incrément.

⁸Sans chercher à défendre l'idée qu'il s'agit là d'une explication de la loi de Weber-Fechner : il est très vraisemblable que cette explication réside au moins pour partie dans des problématiques de capacités de discriminations affectées par la variabilité des informations sensorielles. De fait, l'approche sensorimotrice est compatible avec n'importe quel seuil différentiel de détection.

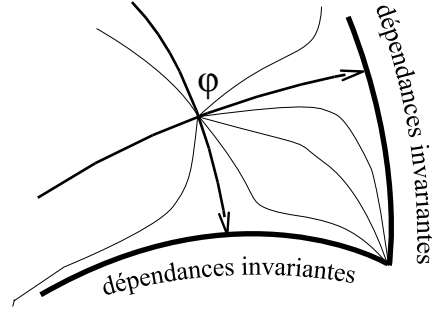


FIG. 6.4 – Limites d'itérations d'un même changement compensable et modalités sensorielles. Les points du plan représentent les dépendances φ' tel que le changement (φ, φ') est compensable par une adaptation motrice non spatiale, les courbes représentent les évolutions de dépendance compatibles avec une même adaptation motrice. Les changements de situation perceptive comportant à la fois un changement de luminosité ou un changement d'intensité sonores soit divergent (au moins l'un des changements est une augmentation), soit convergent vers une même situation (les deux changements sont une diminution). Les changements de situation limités à un changement de luminosité ou un changement d'intensité sonores sont dans ce contexte des évolutions bien particulières.

6.5.8 Régularité, formulation différentielle

L'exemple considéré possède également certaines propriétés de régularité : les ensembles Φ et \mathcal{H} peuvent être dotés d'une structure de variété différentiable. Il est donc possible de traduire dans un cadre différentiel l'ensemble des points précédents :

- on s'intéresse alors aux changements tangents de dépendance et d'adaptation motrices : $(\varphi, \dot{\varphi}) \in T_{\varphi}\Phi$ et $\dot{h} \in T_{id_U}\mathcal{H}$,
- on dira que $(\varphi, \dot{\varphi})$ est un changement compensable par l'adaptation motrice \dot{h} s'il existe des trajectoires régulières de dépendances φ_t et d'adaptations motrices h_t telles que (φ, φ_t) est compensable par h_t pour tout t , avec : $\varphi_0 = \varphi$, $h_0 = id_U$, et les vecteurs tangents en $t = 0$ des trajectoire φ_t et h_t sont $\dot{\varphi}$ et \dot{h} ,
- on dira que la trajectoire φ_t est compensable par \dot{h} si $(\varphi_t, \dot{\varphi}_t)$ est compensable par \dot{h} pour tout t ,
- on dira que \dot{h} est une adaptation motrice génériquement compensable si pour tout $\varphi \in \Phi$, il existe $(\varphi, \dot{\varphi})$ compensable par \dot{h} ,
- on dira que la dépendance φ a des invariances si $(\varphi, 0)$ est compensable par une famille d'adaptations motrices \dot{h} ,
- il existe un système de correspondances $\dot{h}_1 \dot{h}_2 \sim \dot{g}$ sur les éléments de $T_{id_U}\mathcal{H}$ si

la succession d'une trajectoire compensable par \dot{h}_1 et d'une trajectoire compensable par \dot{h}_2 , produit une trajectoire compensable par \dot{g} et ainsi de suite.

L'existence, dans l'exemple considéré, de trajectoires régulières de dépendances φ_t et d'adaptations motrices h_t telles que (φ, φ_t) est compensable par h_t pour tout t est un fait qui n'est pas anodin. Le champ réceptif étendu des photorécepteurs exprime précisément que l'activité neuronale sensorielle dépend de manière régulière de l'environnement et de la posture du dispositif visuel, dans le cas d'un environnement constitué par une source lumineuse ponctuelle⁹. L'étendue du champ réceptif des photorécepteurs est la conséquence de différents facteurs, entre autre : extension physique du récepteur et filtrage spatial de l'image rétinienne par l'optique oculaire.

⁹Cela implique la régularité de l'activité sensorielle dans le cas d'un environnement constitué par une superposition de sources ponctuelles, du fait de la linéarité des équations de Maxwell et de l'additivité de l'absorption des photons au niveau réceptoral.

Chapitre 7

L'argumentation de Poincaré

Ce qui suit est un extrait de l'argumentation de Poincaré sur l'origine de la géométrie, à partir de la notion de compensabilité (Poincaré, 1898), suivi d'une analyse des difficultés d'interprétation de cette argumentation.

7.1 Des fondements de la géométrie

DÉPLACEMENT ET CHANGEMENT D'ÉTAT

Nous percevons du premier coup que nos sensations varient, que nos impressions sont sujettes au changement. Les lois de ces variations ont été cause que nous avons créé la géométrie et la notion d'espace géométrique. Si nos sensations n'étaient pas variables, il n'y aurait pas de géométrie.

Mais ce n'est pas tout. La géométrie ne serait pas née si nous n'avions pas été amenés à répartir en deux classes les changements que peuvent subir nos impressions. Nous disons tantôt que nos impressions ont changé parce que les objets qui les produisaient ont subi quelque changement d'état et tantôt que nos impressions ont changé parce que les objets ont subis un déplacement. Quel est le fondement de cette distinction ?

Une sphère dont un hémisphère est bleu et l'autre rouge tourne sur elle-même devant nos yeux et montre d'abord un hémisphère bleu et ensuite un hémisphère rouge. Un liquide bleu contenu dans un vase subit une réaction chimique qui le fait devenir rouge. Dans les deux cas l'impression de bleu a été remplacée par celle de rouge. Pourquoi le premier de ces changements est-il classé parmi les déplacements et le second parmi les changements d'état ? Evidemment parce que dans le premier cas il me suffit de tourner autour de la sphère pour me placer de nouveau vis-à-vis de l'autre hémisphère et éprouver ainsi une seconde fois l'impression de bleu.

Un objet se déplace devant mon oeil et son image qui se formait d'abord au centre de ma rétine se forme maintenant sur le bord. La sensation qui m'était apportée par une fibre nerveuse partant du centre de la rétine est remplacée par une autre sensation qui m'est apportée par une fibre partant du bord de la rétine. Ces sensations me sont amenées par deux nerfs différents. Elles doivent m'apparaître comme qualitativement différentes, et, s'il n'en était pas ainsi, comment pourrais-je les distinguer ?

Pourquoi alors suis-je amené à penser que c'est la même image qui s'est déplacée ? Est-ce parce que l'une de ces sensations succède souvent à l'autre ? Mais des successions analogues sont fréquentes. Ce sont elles qui font naître toutes nos associations d'idées et nous n'en concluons ordinairement pas qu'elles sont dues au déplacement d'un objet qualitativement invariable.

Mais ce qui arrive dans ce cas, c'est que nous pouvons suivre l'objet de l'oeil et, par un déplacement de l'oeil qui est généralement volontaire et accompagné de sensations musculaires, ramener l'image au centre de la rétine et rétablir ainsi la sensation primitive.

Je conclurai donc comme il suit :

Parmi les changements qui subissent nos impressions, nous distinguons deux classes :

- 1. Les premiers sont indépendants de notre volonté et ne sont pas accompagnés de sensations musculaires. Ce sont des changements externes, pour ainsi dire.*
- 2. Les autres sont volontaires et accompagnés de sensations musculaires. Nous pouvons les appeler changements internes.*

Nous observons ensuite que dans certains cas, lorsqu'un changement externe a modifié nos impressions, nous pouvons, en provoquant volontairement un changement interne, rétablir nos impressions primitives. Le changement externe peut donc être corrigé par un changement interne. Les changements externes peuvent être, par conséquent, subdivisés en deux classes :

- 1. Les changements qui sont susceptibles d'être corrigés par un changement interne. Ce sont les déplacements.*
- 2. Les changements qui n'en sont pas susceptibles. Ce sont les changements d'état.*

Un être qui ne pourrait pas se mouvoir serait incapable de faire cette distinction. Un tel être, par conséquent, ne pourrait jamais créer la géométrie, - même si ses sensations étaient variables et même si les objets qui l'entourent étaient mobiles.

CLASSIFICATION DES DÉPLACEMENTS

Une sphère dont un hémisphère est bleu et l'autre rouge tourne sur elle-même devant moi et me présente d'abord son côté bleu et ensuite son côté rouge. Je regarde ce changement externe comme un déplacement parce que je peux le corriger par un changement interne, c'est-à-dire en

tournant autour de la sphère. Répétons l'expérience avec une autre sphère dont un hémisphère est vert et l'autre jaune. L'impression de l'hémisphère jaune succédera à celle du vert comme auparavant celle du rouge succédait à celle du bleu. Pour la même raison je regarderai ce nouveau changement externe comme un déplacement.

Mais ce n'est pas tout. Je dis aussi que ces deux changements externes sont dus au même déplacement, c'est-à-dire à une rotation. Cependant, il n'y a pas de rapport entre l'impression de l'hémisphère jaune et celle du rouge, pas plus qu'il n'y en a entre celle de l'hémisphère bleu et celle du vert, et je n'ai aucune raison de dire qu'il existe entre le jaune et le vert la même relation qu'entre le rouge et le bleu. Non, je dis que ces deux changements externes sont dus au même déplacement parce que je les ai corrigés par le même changement interne. Mais comment puis-je savoir que les deux changements internes par lesquels j'ai corrigé d'abord le changement externe du bleu au rouge, et ensuite celui du vert au jaune, doivent être considérés comme identiques ? Tout simplement parce qu'ils ont provoqué les mêmes sensations musculaires ; et pour cela je n'ai pas besoin de connaître d'avance la géométrie et de me représenter les mouvements de mon corps dans l'espace géométrique.

Ainsi plusieurs changements externes qui n'ont en eux-mêmes rien de commun peuvent être corrigés par le même changement interne. Je les rassemble dans la même classe et les considère comme le même déplacement.

Une classification analogue peut être faite en ce qui concerne les changements internes. Tous les changements internes ne sont pas capables de corriger un changement externe. Seuls ceux qui en sont capables peuvent être appelés déplacements. D'autre part, le même changement externe peut être corrigé par différents changements internes. Une personne qui saurait la géométrie pourrait exprimer cette idée en disant que mon corps peut aller de la position A à la position B par plusieurs chemins différents. Chacun de ces chemins correspond à une série de sensations musculaires et pour le moment je n'ai pas connaissance d'autre chose que de ces sensations musculaires. Il n'y a pas de ressemblance entre deux de ces séries et si je les considère néanmoins comme représentant le même déplacement, c'est parce qu'elles sont susceptibles de corriger le même changement externe.

La classification qui précède suggère deux réflexions :

- 1. La classification n'est pas une donnée brute de l'expérience, parce que la compensation mentionnée plus haut des deux changements, l'un interne et l'autre externe, n'est jamais exactement réalisée. C'est donc une opération active de l'esprit qui essaie d'insérer les résultats bruts de l'expérience dans une forme préexistante, dans une catégorie. Cette opération consiste à identifier deux changements parce qu'ils possèdent un caractère commun, et cela malgré qu'ils ne le possèdent pas exactement. Néanmoins, le fait même que l'esprit ait l'occasion d'accomplir cette opération est dû à l'expérience, car*

l'expérience seule peut lui apprendre que la compensation s'est approximativement produite.

2. *La classification nous amène en outre à reconnaître l'identité de deux déplacements et il en résulte qu'un déplacement peut être répété deux ou plusieurs fois. C'est cette circonstance qui introduit le nombre et permet la mesure là où régnait auparavant la pure qualité.*

INTRODUCTION DE LA NOTION DE GROUPE

Que nous soyons capables d'aller plus loin est dû au fait suivant dont l'importance est capitale.

Il est évident que si nous considérons un changement A et le faisons suivre d'un autre changement B, nous sommes libres de regarder l'ensemble des deux changements A suivi de B comme un seul changement qui peut s'écrire $A + B$ est peut être appelé le changement résultant. (Il va sans dire que $A + B$ n'est pas nécessairement identique à $B + A$). Il en résulte alors que si deux changements A et B sont des déplacements, le changement $A+B$ est aussi un déplacement. Les mathématiciens expriment cela en disant que l'ensemble des déplacements forme un groupe. S'il n'en était pas ainsi il n'y aurait pas de géométrie.

Mais comment savons-nous que l'ensemble des déplacements est un groupe ? Est-ce par un raisonnement a priori ? Est-ce par expérience ? On est tenté de raisonner a priori et de dire : si le changement externe A est corrigé par le changement interne A', et le changement externe B par le changement interne B', le changement externe résultant $A+B$ sera corrigé par le changement interne résultant $B' + A'$. Donc ce changement résultant est par définition un déplacement, ce qui revient à dire que l'ensemble des déplacements forme un groupe.

Mais ce raisonnement est sujet à plusieurs objections. Il est clair que les changements A et A' se compensent, c'est-à-dire que si ces deux changements se succèdent, je retrouverai mes impressions primitives, résultat que je peux écrire comme il suit : $A + A' = O$.

Je vois également que $B + B' = O$. Ce sont ces hypothèses que j'ai faites au début et qui m'ont servi à définir les changements A, A', B et B'. Mais est-il certain que nous aurons encore $B + B' = O$ après les deux changements A et A' ? Est-il certain que ces deux premiers changements se compensent d'une manière telle qu'après eux, non seulement je retrouverai mes impressions primitives, mais que les deux changements B et B' retrouveront toutes leurs propriétés initiales et en particulier celle de se compenser mutuellement ? Si nous admettons cela, nous pourrions en conclure que je retrouverai mes impressions primitives quand les quatre changements se suivront dans l'ordre A, A', B, B' mais non pas qu'il en sera encore de même quand ils se succéderont dans l'ordre A, B, B', A'.

Et ce n'est pas tout. Si deux changements externes a et a' sont regardés comme identiques

sur la base de la convention adoptée plus haut, ou, en d'autres termes, sont susceptibles d'être corrigés par le même changement interne A ; si, d'autre part, deux autres changements externes b et b' peuvent être corrigés par le même changement interne B et peuvent par conséquent être regardés aussi comme identiques, avons-nous le droit de conclure que les deux changements $a + b$ et $a' + b'$ sont susceptibles d'être corrigés par le même changement interne et sont par conséquent identiques ? Une telle proposition n'est aucunement évidente, et, si elle est vraie, elle ne peut être le résultat d'un raisonnement a priori.

Par conséquent, cette série de propositions, que je résume en disant que les déplacements forment un groupe, ne nous est pas donnée par un raisonnement a priori. Sont-elles donc un résultat d'expérience ? On est enclin à admettre qu'elles le sont ; et cependant on a un sentiment de véritable répugnance à le faire. Des expériences plus précises ne peuvent-elles pas prouver un jour que la loi énoncée plus haut n'est qu'approximative ? Et alors qu'advient-il de la géométrie ?

Mais nous pouvons être tranquilles sur ce point. La géométrie est à l'abri de toute révision ; aucune expérience, si précise soit-elle, ne peut la renverser. Si cela se pouvait, il y a longtemps que se serait fait. Nous savons depuis longtemps que toutes les soi-disant lois expérimentales ne sont que des approximations et des approximations grossières.

Que faut-il donc faire ? Quand l'expérience nous apprend qu'un certain phénomène ne correspond pas du tout aux lois indiquées, nous l'effaçons de la liste des déplacements. Quand elle nous apprend qu'un certain changement ne leur obéit qu'approximativement, nous considérons ce changement, par une convention artificielle, comme la résultante de deux autres changements composants. Le premier composant est regardé comme un déplacement satisfaisant rigoureusement aux lois dont je viens de parler, tandis que le second composant, qui est petit, est regardé comme une altération qualitative. Ainsi nous disons que les solides naturels ne subissent pas seulement de grands changements de position, mais aussi de petites flexions et de petites dilatactions thermiques.

Par un changement externe a par exemple, nous passons de l'ensemble d'impressions A à l'ensemble B. Nous corrigeons ce changement par un changement interne volontaire b et nous sommes ramenés à l'ensemble A. Un nouveau changement externe a' nous fait passer de nouveau de l'ensemble A à l'ensemble B. Nous devons nous attendre alors à ce que ce changement a' puisse à son tour être corrigé par un autre changement interne volontaire b' qui provoquerait les mêmes sensations musculaires que b et qui ramènerait l'ensemble d'impressions A. Si l'expérience ne confirme pas cette prédiction, nous ne sommes pas embarrassés. Nous disons que le changement a' , bien qu'il nous ait fait, comme a , passer de l'ensemble A à l'ensemble B, n'est cependant pas identique au changement a . Si notre prédiction ne se confirme qu'approximativement, nous disons que le changement a' est un déplacement identique au déplacement a , mais

accompagné d'une légère altération qualitative.

En résumé, les lois en question ne nous sont pas imposées par la nature, mais sont imposées par nous à la nature. Mais si nous les imposons à la nature, c'est parce qu'elle nous permet de le faire. Si elle offrait trop de résistance, nous chercherions dans notre arsenal une autre forme qui serait pour elle plus acceptable.

7.2 Les difficultés d'interprétation de l'argumentation

Le travail épistémologique de Poincaré sur l'origine de la géométrie est écrit en langage courant. Sa proposition est ostensiblement appuyée sur des concepts mathématiques établis, mais elle reste, de manière déroutante, au stade d'une narration. Sans doute, Poincaré devait juger son discours assez clair pour pouvoir être traduit immédiatement en langage mathématique par qui le souhaiterait. Car même si, comme il le répète en plusieurs endroits, son histoire repose sur une conception idéalisée de la situation, il doit bien être possible de déployer une théorie mathématique reflétant cette conception idéalisée, à défaut de déployer une théorie mathématique reflétant la situation réelle.

Le problème du langage courant employé par Poincaré est que, en plusieurs endroits critiques, son discours paraît ambigu. La notion de sensations musculaires qui informe son géomètre empiriste semble en particulier renvoyer tantôt aux changements de posture, tantôt aux changements de position de son organisme. Cela induit une confusion problématique.

Il ne s'agit pas de rentrer dans le détail de l'argumentation de Poincaré, mais de revenir sur quelques points importants, pour montrer la difficulté de son interprétation. J'utilise dans cette section les termes utilisés par Poincaré, afin d'être en ligne avec ses citations, mais cela rentre en conflit avec la position que j'ai exprimé dans la première partie. Poincaré prend en effet dans son argumentation une attitude phénoménaliste : il réfère à des "impressions" (visuelles) et des "sensations" (musculaires) comme à "des choses entrant dans la conscience", y étant présentes ou pas.

Tout d'abord, Poincaré ne précise nulle part ce qu'il a en tête par l'expression "changement d'impressions". Il semble raisonnable d'imaginer qu'il s'agit de couples d'impressions (s_1, s_2) , mais il pourrait s'agir également d'un concept élémentaire. Ensuite, lorsqu'il distingue, parmi les changements d'impression, les changements ex-

ternes et internes, Poincaré écrit :

1. Les premiers [...] ne sont pas accompagnés de sensations musculaires. Ce sont des changements externes [...]
2. Les autres sont [...] accompagnés de sensations musculaires. Nous pouvons les appeler changements internes.

mais il ne précise pas ce qu'il entend par "être accompagné ou pas de sensation musculaire". Or on peut imaginer deux solutions. Soit Poincaré considère que les données de son géomètre empiriste sont de la forme (s_1, s_2, m) , où m est une sensation musculaire qui peut prendre une valeur distinguée signifiant l'absence de sensation. Soit qu'elles sont de la forme (s_1, s_2, m_1, m_2) , et par absence de sensation musculaire il veut en fait dire absence de changement de sensation musculaire, i.e. $m_1 = m_2$.

Les deux solutions ne sont pas identiques, elles dotent le sujet de capacités différentes : dans le second cas, si $m \neq m'$, il est possible de distinguer des situations (s_1, s_2, m, m) et (s_1, s_2, m', m') , qui ne peuvent être distinguées dans le premier cas. La première solution suggère que "la conscience" considérée par Poincaré n'a pas accès à la posture du corps (sinon, qu'est-ce que l'absence de sensation) mais seulement à ses changements, l'autre suggère que nous l'avons, au moins en partie : l'une des solutions suggère une dépendance associative, l'autre une dépendance différentielle.

Ces difficultés sont suivies par une autre au moment, pourtant critique, de la discussion sur la composition :

Il est évident que si nous considérons un changement A et le faisons suivre d'un autre changement B , nous sommes libres de regarder l'ensemble des deux changements A suivi de B comme un seul changement qui peut s'écrire $A + B$ est peut être appelé le changement résultant.

Paradoxalement, alors que bien sûr c'est dans la définition d'une loi de composition que se situe le coeur du problème, Poincaré se contente de dire qu'il est évident qu'on peut la définir. Clairement, il ne peut pas être en train de parler de l'enchaînement de deux changements d'impression (s_1, s_2) et (s_2, s_3) pour former (s_1, s_3) , puisqu'il ne serait pas possible de "faire suivre" n'importe quels changements, par exemple les changements (s_1, s_2) et (s_3, s_4) .

Poincaré pourrait utiliser les lettres A et B pour référer à des classes de changements d'impressions internes associées à une même sensations musculaire, mais il n'y

aurait alors rien d'évident au fait qu'il soit possible de trouver dans A et B des éléments de la forme (s_1, s_2) et (s_2, s_3) pour pouvoir associer à (A, B) la classe de (s_1, s_3) , ni d'ailleurs qu'il fait que ce dernier changement soit associé à un changement interne, donc à une nouvelle classe C .

Il n'y a rien "d'évident", mais on pourrait considérer qu'il s'agit de faits qui doivent se trouver vérifiés de manière empirique. Est-il raisonnable, cependant, de l'espérer ? Il me semble que Poincaré suppose qu'il suffit de piocher un changement (s_1, s_2) dans A , et qu'il existe ensuite forcément un changement de la forme (s_2, s_3) dans B . Or imaginons que B est l'ensemble des changements d'impression associés à un même changement (m_1, m_2) , que m_1 corresponde à une posture où les yeux louchent et que s_2 corresponde aux images rétinienne en vision normale. On ne sera jamais confronté en même temps à s_2 et m_1 , il n'y aura donc pas d'élément (s_2, s_3) dans B .

Enfin, Poincaré ne peut être en train de parler de la composition des déplacements, la seule pourtant qui me semble véritablement possible à définir, puisque la phrase suivante du texte présente les déplacements comme un cas particulier de cette composition plus généralement des changements.

Le reste du raisonnement, enfin, me semble de nouveau très difficile à suivre réellement du fait de l'ambiguïté déjà évoquée, et due au langage courant utilisé par Poincaré, sur la référence à une dépendance associative ou différentielle.

Chapitre 8

Articles en préparation : dépendances différentielles

Cette annexe rassemble deux articles non finalisés. Ils présentent des travaux de recherche sur une évolution des techniques de réduction de dimension dans un contexte sensorimoteur. Alors que les techniques de réduction de dimension classiques s'appuient sur une notion de distance entre les données qui est déterminée par la représentation de ces données, l'idée était de proposer d'abord une étape d'apprentissage des dépendances sensorimotrices caractéristiques du dispositif d'acquisition des données, pour déterminer une notion de distance reflétant la structure de ces dépendances.

Le premier article¹ se base sur l'apprentissage d'un ensemble continu de transformations des informations sensorielles, transformations associées à différents mouvements. L'estimation de leur structure de composition permet alors d'extraire un ensemble de transformations élémentaires, utilisées pour mettre en relations les différentes informations sensorielles. Cet article était accepté par Neural Computation sous réserves de modifications, qui n'ont toujours pas été faites faute de temps.

Le second article, encore à l'état d'ébauche, reprend la base du précédent, en cherchant à contourner la difficulté que représente l'apprentissage d'un ensemble de transformations des informations sensorielles et l'analyse de leur structure de composition. L'idée est de proposer à la place une phase d'expérimentation sensorimotrice conduisant à distinguer quelques mouvements élémentaires, et de se limiter à l'apprentissage

¹Sans en être le premier auteur, j'ai contribué à cet article de manière très active, tant au niveau de la définition de ses objectifs, que de la méthode employée, que de l'analyse et l'argumentation générale.

de détecteurs de modifications sensorielles compatibles avec ces mouvements.

Sensorimotor Reduction of Dimensionality

Frédéric Glanois, David Philipona, J.Kevin O'Regan
Laboratoire de Psychologie de la Perception, CNRS
Université René Descartes, Paris V

Abstract

Numerous AI domains are concerned with nonlinear dimensionality reduction. Whereas the metric used in classical methods is usually derived from the sensory input data itself, we claim that in the case of an embodied agent, a more natural metric arises, provided by the intrinsic structure of the agent's capacities to act in the world.

We apply this idea to the particular case where the actions involved correspond to spatial displacements of an agent. We detail an algorithm and demonstrate its performance in a simple simulation where pictures acquired by a moving camera are to be embedded in the three-dimensional manifold generated by the camera's translations and rotations in the plane. The method also allows isolation of the flat subspaces generated by translations alone.

1 Introduction: dimensionality reduction

Living organisms and artificial agents face the world through sensory channels providing a huge amount of tremendously complex information. While the actual way biological organisms deal with this complexity still lies beyond our present understanding, tackling the question cannot be avoided in embodied AI. If an artificial agent is to develop interaction strategies of algorithmic nature, it needs to extract from the perceptual data some sensible structure, some regularities that will turn the maelstrom into a workable data set.

In this perspective, dimensionality reduction (hereafter noted DR) consists in exhibiting an accurate low-dimensional description for data initially described in a high-dimensional space, when the data points actually lie on a low-dimensional manifold. Provided with this economical representation, the agent will then be more likely to succeed in tasks such as statistical estimation or planning with limited time and memory resources.

Besides well established linear techniques such as PCA, nonlinear DR approaches have made progress in the last decade. Most of them are formally

described as consisting in unfolding a flat¹ low-dimensional submanifold of a Euclidean space. In the example of the well-known Swiss Roll (see [2]), data points belong to a rolled plane in \mathbb{R}^3 and the reduction algorithm aims at exhibiting a \mathbb{R}^3 -to- \mathbb{R}^2 mapping whose restriction to the roll is an *isometric*² embedding.

Classically, reduction techniques do not care *how* the data have been collected. The metric used to interpret the data and ultimately build the low-dimensional embedding is simply derived from the initial data representation. Yet in the context of embodied systems, the target representation is *meant to afford the agent's interaction with the world*. From this consideration a natural idea follows: one may look for a metric on the sensory inputs that directly relates to the physical abilities of the acting system. In this paper we intend to demonstrate the feasibility and relevance of this novel approach.

After briefly reviewing the state of the art, we evoke some practical issues inherent to “passive” methods. We subsequently explain how an embedded system can, upon experimenting its acting capabilities, discover a natural, intrinsic metric with which to interpret its sensory inputs. Then we present an algorithm by which a sensorimotor system learns to structure its abilities, extracts the corresponding code-invariant metric, and applies it to perform DR on perceptual data. We detail a simulation in a simple example of 2D images. We finally discuss the background, performance and generality of the method.

2 Passive understanding of the world

2.1 Classical approach

Many unsupervised learning algorithms have been developed in the last decade: **MDS**, **Isomap** [1], **Locally Linear Embedding** [2], **Laplacian Eigenmaps** [3], **Curvilinear Component Analysis** [4], **Kernel PCA** [5] or **Charting** [6] are established standard examples. Bengio and al. [7] cast these methods into a common Kernel framework and calculate the corresponding adaptation formulas. The first step is to compute a “similarity matrix” M such that M_{ij} contains the distance between the data points P_i and P_j . Then a data-dependent kernel can be derived from it, whose ordered eigenvectors and eigenvalues give rise to the embedding through the so-called Nyström formula. See details in [7].

Note that other approaches to DR do not fit in this framework. These

¹A submanifold with no intrinsic curvature

²An isometric mapping preserves the local metric structure of the original space, which ensures that the distance relations existing within the original manifold are not arbitrarily distorted.

include supervised learning methods, graph algorithms, or distributed approaches such as **Self-Organizing Maps** [8]³.

Nevertheless all DR algorithms rely, explicitly or not, on a notion of distance. A similarity matrix implies a claim about what it is for two sensory inputs to be close to each other. Importantly, from the point of view of an embodied agent, the above methods never take into account the actions that generated the sensory inputs. In this respect we term them as *passive*.

2.2 Practical issues of using a passive metric

Take vision, and a simple $n \times n$ picture of vertical alternating black/white stripes; compare it with the same picture translated horizontally by the width of one stripe. Both pictures correspond to a point in \mathbb{R}^{n^2} : how close together are they? If we apply the Euclidean metric, which involves a pixel by pixel computation, the result is maximal, indicating a very high *dissimilarity*. Yet we consider an *embodied* agent getting such data as input: because the two points represent pictures translated by a small amount, we wish them to be considered *close* to each other. In other words, we deem that the agent's notion of similarity should more closely relate to its interaction with the world.

This example is somewhat artificial because in natural images bandwidth and continuity prevent Euclidian distance from appreciably misleading action-oriented processing. Yet a similar difficulty unavoidably arises if the sensor is damaged, has imperfections (as biological systems have), or is badly calibrated. In artificial vision for instance, an image-based metric is bound to classify the inputs based on artifacts that the camera optics and sampling induce.

In short, the embedding exhibited by any passive approach (most of the above techniques rely on the Euclidian metric) seems inappropriate in that it does not come with a clear interpretation with respect to action. Consequently the coordinate system obtained by the reduction might not highlight the important dimensions, as far as the agent's task of interacting with the world is concerned, of the data manifold.

3 Deriving a metric from interaction

3.1 Relating perception to action capabilities

Take the example of a camera which can be moved around by motors. Suppose the system starts by undergoing an exploratory phase. If the camera has d degrees of freedom, it can experiment in the resulting d -dimensional sensorimotor space until it is able to somehow predict the sensory consequences of its actions. Then the agent will have access to a truly relevant

³For a list of references, see <http://www.cse.msu.edu/~lawhiu/manifold/>

metric for the data: two inputs are close *if they can be obtained one from another through a small action*⁴.

More generally, we consider an agent interacting with a world, and we assume it to induce a group of symmetries of the agent-environment system (see 3.2). The idea is to learn the structure of this group and build a corresponding symmetry *basis* - or at least to define constraints for its choice. This basis strongly relates to the acting capabilities of the agent; from there a natural metric can be defined to within arbitrary scaling factors. As explained, we consider such an action-based metric to be a relevant choice for DR. See section 5 for the description of our algorithm.

3.2 Mathematical sketch of the approach

We start by explaining the building of the metric, then we dwell on the extraction of flat submanifolds; finally we describe the hypotheses of the model.

3.2.1 The metric

Let Ω be the manifold of a sensorimotor organism's states. We assume a law $\phi(\cdot, m)$ describing the consequences of a given motor command $m \in \mathcal{M}$:

$$\dot{\omega} = \phi(\omega, m) \quad \forall \omega \in \Omega \quad (1)$$

The crucial point is that we assume the existence of symmetries for the system: regularities underlying the interaction of the agent with the world. In other words, we assume a set \mathfrak{T} of differentiable mappings $T : \Omega \rightarrow \Omega$ whose action leaves the above equation invariant:

$$\forall T, \quad \phi(T(\omega), m) = DT(\omega) \cdot \dot{\omega} \quad (2)$$

where DT is the linear mapping associated with the symmetry T , which transports the tangent space around ω when T transports ω . Picture a moving agent in a Euclidian space: the *relationships* between the actions it can perform do not change if carried to another spot (the Euclidian displacements are symmetries of the system).

As we aim at building a metric⁵ relying on Eq. 1, we want this metric to remain invariant under the same symmetries. More technically, a metric $g(\cdot, \cdot)$ defines a scalar product in every point and thus we demand that for any tangent vectors $(\dot{\omega}_1, \dot{\omega}_2)$ around ω :

$$g_\omega(\dot{\omega}_1, \dot{\omega}_2) = g_{T(\omega)}(DT(\omega)\dot{\omega}_1, DT(\omega)\dot{\omega}_2) \quad \forall \omega \in \Omega \quad (3)$$

⁴The notion of “small” action will be clarified later, but can be understood intuitively for the moment.

⁵Mathematically, a metric is a differentiable mapping defining, for each point p of the manifold, a scalar product on $T_p\Omega$: $g : p \mapsto g_p(\cdot, \cdot)$

If we assume that the set of actions allows the whole state space to be explored ⁶, it is then straightforward to verify:

$$g_w(\phi(m_1, \omega), \phi(m_2, \omega)) = \text{constant} \quad \forall \omega \quad (4)$$

This equation expresses what we call the *land survey lemma*:

Land Survey Lemma. *Given two motor commands, the scalar product between the resulting tangent vectors is constant across Ω .*

Thus, in order to define the metric on the whole space, it is sufficient to choose a set of commands $\{m_i\}$ such that $\{\phi(m_i, \omega)\}$ is a basis of the tangent space at ω . By symmetry, it is a basis everywhere, and a canonical metric is derived.

3.2.2 Flatness and curvature

It is generally relevant to the goal of DR to separate subspaces that possess an intrinsic curvature (and thus cannot be properly embedded in an Euclidian space) from subspaces that do not. As Fig. 1 illustrates, this relates to commutativity properties of the vector fields. Focusing on the study of a moving agent in a 2D world (see later), we look for the particular subset of symmetries which commute between each other: these are the translations \mathcal{T} . Using our knowledge of the group structure of $SE(2)$, we specify a core constraint on the basis: it should be of the form $\{T_1, T_2, M\}$ where $M \notin \mathcal{T}$.

It should be stated clearly that, while the algorithm is general, as well as invariant with respect to the encoding of sensorimotor information, it was not written in a general way, but adapted to the case of $SE(2)$.

Note that, *as far as representing the data in a low-dimensional Euclidian space is concerned*, it is necessary to restrict the data set to a subspace where curvature can be discarded, unless we just want to represent only the two-dimensional flat submanifold.

3.2.3 Hypotheses of the model

The agent does not know about its state, but has access to a sensory input $x = \varphi(\omega)$; consequently the metric has to be built on $\mathcal{X} = \varphi(\Omega)$. Herein we assume that each action induces unambiguously distinguishable sensory modifications. Given two neighbor states, there has to be enough information in the sensors to recognize the “tangent” action that has led from one underlying state to the other.

We replace Eq. 1 in the following way: we assume an linear approximation for the sensorimotor dynamics of the system. More precisely, each

⁶Mathematically, we say that the action of the group \mathfrak{T} is *transitive*.

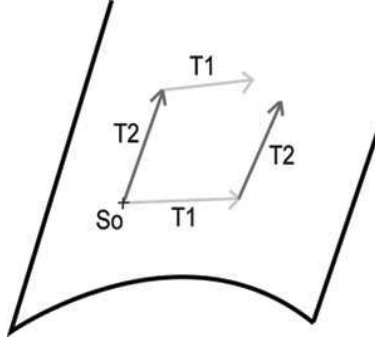


Figure 1: The curvature of a manifold expresses the non-closure of a parallelogram: in short, the operations of parallel transport along autoparallel curves do not commute.

action m is assumed to induce a transformation of the sensory input X of the form:

$$\dot{X} = P^m \cdot X + \eta^m \quad (5)$$

We will refer to the mapping P^m as the linear *predictor* of the action m . “Mastering” m amounts to learning P^m and η^m . The latter expresses the intrinsic indeterminacy of the prediction: if a camera is shifted, only the pixels that stay on the screen can be exactly predicted, while the border ones are variable. We discard the directions in which there is too much variability (the individual sensors with highest variance): as long as we have sufficient remaining “reliability subspaces” for each predictor, we can compare and recognize these. We also avoid large motions and composition of actions.

4 Related work

Close in spirit, Philipona *et al.* [17] demonstrate how a sensorimotor organism can discover the geometry of the space it is embedded in. They provide an algorithm extracting the structure of the sensorimotor interaction using Lie brackets, and demonstrate its practicability in a toy simulation. But their agent is not concerned with sensory prediction. Moreover, the environment is conveniently assumed to allow transport of structure information across the manifold.

Rao and Ruderman [18] propose an algorithm estimating the generator matrix G of a Lie group; their work is extended by Bengio and Monperrus [19] who develop a method for non-local manifold learning. Their unsupervised algorithm learns and predicts the transformation on which it is trained, but as it essentially deals with one dimension, it does not really face a structure to extract. The concern of growing indeterminacy hampering large-scale predictions is solved by a cyclic hypothesis.

In recent work oriented towards MRI, Pless [20] and Pless and Souvenir [21] reckon that traditional DR methods produce parameters that “do not have explicit meaning”; given pictures of an object undergoing unknown deformations, they aim at tracing back the journey. They propose to introduce an appropriate distance measure in **Isomap** using complex Gabor filters applied on image samples. Another approach is Simard *et al.*’s [22] Tangent Distance, “a modification of the standard distance measure to allow it to incorporate a priori knowledge specific to the problem”. Basically, they study the manifold generated around a perceptual point by linear spatial transformations. Although sharing important concerns, both approaches differ from ours because they do not demand that the agent acquire its knowledge through actual interaction with the world.

5 The algorithm

5.1 Outline of the method

Our proposal amounts to adding a learning phase, which obviously occurs just once, to classic DR methods.

Learning of the sensorimotor laws :

- Experiment the interaction with the world, learn to *predict* the sensory consequences of a given action in a given situation.
- Build an *action basis* reflecting the intrinsic structure of the agent’s abilities; a natural metric results.

Dealing with the data :

- Using the new metric, compute the “active” distance between neighboring points
- Fill the similarity matrix using Floyd’s algorithm
- Go on as **MDS**.

5.2 Detailed description of the learning phase

Here we describe the example of simulation we chose and detail the actual implementation of the algorithm.

5.2.1 The world

We consider a 2D world (a large gray-scale image). The sensorimotor device is constituted by a moving camera, whose size is 21×21 pixels, and whose input is multiplied by some distortion matrix unknown to the system. The device can displace itself on the image: each move is implemented

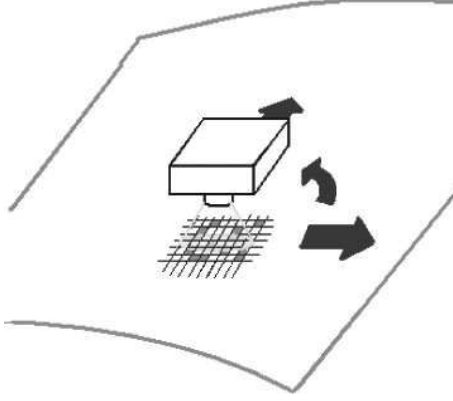


Figure 2: The camera roaming the 2D world

as a transformation $\{\delta x, \delta y, \delta \theta\}$ of the spatial configuration $\{x, y, \theta\}$. These transformation are restricted to a small interval $([.6, 1.4] \times [.6, 1.4] \times [-12, 12]$ in pixels \times pixels \times degrees) but chosen randomly .

5.2.2 Learning the predictors

The system starts by experimenting in turn a certain number of random moves ($N = 40$) which are neither pure translations nor rotations, plus their N^2 combinations. For each action and combination, a linear regression is performed over a sufficient number of pairs of {after/before} sensory inputs to learn the predictor⁷ P^θ or $P^{\theta\mu}$ (see Eq. 5).

The next step is to extract a relevant *basis*: a small set of predictors such that any small motion of the system can be accurately approximated by a linear combination of those. As explained in subsection 3.2.2, a relevant basis for the case pictured in Fig. 2 can be decomposed into a two-dimensional commutative subgroup (the translations) and a complementary one-dimensional subgroup (the “rotations”).

5.2.3 The predictor space

Predictors are high-dimensional matrices but lie on a three-dimensional manifold \mathcal{Q} , which cannot a fortiori be approximated by a linear space. Yet we assume that \mathcal{Q} is “smooth enough” to be piecewise linearly approximated. Thus, we can create valid virtual predictors (“virtual” meaning hereafter that the agent has never actually experienced the corresponding action) by linearly combining neighboring ones. We can choose a list of known predictors $\{P^{\mu_i}\}$ and form combinations:

⁷Which is associated either with the index of the action or with the couple of indexes corresponding to the couple of actions.

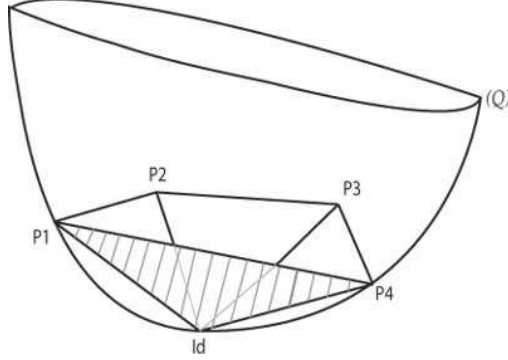


Figure 3: Schema of a two-dimensional predictor space in \mathbb{R}^3 . The surface around the identity is divided into four quadrants delimited by the four predictors $\{P^1, P^2, P^3, P^4\}$. One given predictor in this region can be accurately expressed as a combination of two out of the four elements of the basis.

$$\tilde{P} = \sum_{i=1}^k \alpha_i P^{\mu_i} \quad (6)$$

if and only if the P^{μ_i} are close to each other in the predictor space (in the sense of the Euclidian norm on matrices) and if the weights $\alpha_i \in [0, 1]$. In short, the system builds a piecewise linear paving from the restricted number of motions it knows initially. In the simulation we fixed $k = 4$, so as to explore “volumes” of the three-dimensional predictor manifold.

In order to build a vector basis for the local space around the identity, we assume that the use of *two* predictors for each dimension (corresponding to one motion and its inverse) will do: this subpart of \mathcal{Q} is approximated by the $2^3 = 8$ linear spaces generated by the three couples of predictors. Fig. 3 depicts the simpler situation of a two-dimensional \mathcal{Q} (2×2 predictors) embedded in \mathbb{R}^3 . To obtain meaningful coordinates when projecting on this redundant basis, we add a positivity constraint⁸ using MatLab’s routine LSQNONEG.

5.2.4 Discovering translations

The algorithm randomly chooses two distant motions $\{b, c\}$ and explores virtual predictors in the vicinities of P^b and P^c , say \tilde{P}^β and \tilde{P}^μ , trying to minimize their “commutative factor”:

$$\Delta = \frac{\|\tilde{P}^{\theta\mu} - \tilde{P}^{\mu\theta}\|}{(\|\tilde{P}^\theta\| \|\tilde{P}^\mu\|)} \quad \text{where} \quad \begin{cases} \tilde{P}^\theta = \sum_{i=1}^k \alpha_i P^{\theta_i} \\ \tilde{P}^\mu = \sum_{j=1}^k \beta_j P^{\mu_j} \\ \tilde{P}^{\theta\mu} = \sum_{i,j=1}^k \alpha_i \beta_j P^{\theta_i \mu_j} \end{cases} \quad (7)$$

⁸This leads to points with half of their coordinates equal to zero.

We make use here of the N^2 composed predictors P^{ab} we have learned in the first phase, in order to avoid composition of predictions. ϵ being fixed to a small value, $\Delta < \epsilon$ means we have built two promising candidates for translations. A similar test is applied to prune and order the whole list.

Out of this list of hand-made translations, the algorithm extracts the two first vectors of the basis. We choose them “the furthest apart” in the predictor space, which approximates a notion of orthogonality. Then for each of them the algorithm searches for one more predictor that is approximately the predictor of the inverse action. Remember that one *cannot* trivially build the predictor of the inverse t of a move m from the predictor P^m . What we do is look for predictors P^t that minimize:

$$\|(I + P^m) \circ (I + P^t) - I\|_{Eucl.}^2 = \|P^m + P^t + P^m P^t\|_{Eucl.}^2.$$

The last step is normalisation. Along each “predictor axis” in \mathcal{Q} the scaling is left unconstrained by the theory; we choose a simple Euclidian scaling, which is experimentally validated by the approximate squareness of our plots (Fig 4, 5).

This whole subpart of the algorithm succeeds if enough virtual translations have been learnt and if not, the system looks for other ones.

5.2.5 Completing the basis

Finally the algorithm completes the basis with two rotations. As algebraic properties merely tells commutative subspaces apart from the rest, we choose to search for the one original motion that “commutes the least” (e.g., maximizing Δ in Eq. 7), on average, with the translations of the basis. Then the system uses it to learn its inverse⁹.

Finally the learning phase is over and has produced a structural basis of the system’s physical abilities - in this case six predictors: two translations, their two inverses, plus one rotation and its inverse.

5.3 The reduction phase

The system is presented with the data points. It computes the distance between each of them and its neighbors in the following way. If $\{P^m\}$ are the six predictors of the basis, X_i the point, and X_j the neighbor tested¹⁰, the system uses constrained least squares to find:

$$\alpha = \arg \min_{\alpha \in [0,1]^6} (\| \sum (\alpha_m P^m X_i) - X_j \|^2)$$

⁹Once a predictor $P : X \rightarrow X'$ is learnt, the inverse predictor Q cannot be built, but it can easily be learnt, since $Q : X' \rightarrow X$

¹⁰Note that a large amount of computation is spared by using the Euclidian distance for a first gross neighbor selection.

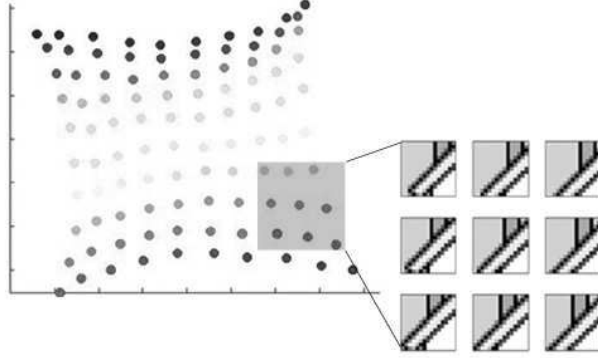


Figure 4: The data is constituted with a grid of translated patches, whose coordinates in the 2D-world are of the form $\{x_0+n, y_0+m\}$ with n and m positive integers. x_0 and y_0 are chosen randomly. In this example the camera does not rotate. *Right*, a sample of these data is shown. 21×21 patches shown are related to each other by one-pixel displacements. *Left*, the output: the embedding of the data points, which is two-dimensional since the only existing dimensions are vertical and horizontal translations. Our algorithm successfully exhibits an approximatively square grid. The curving is a border effect, due to the small quantity of data. Note that there is no reason that the axes correspond to “real” vertical and horizontal axes.

and the distance is taken as $\|\alpha\|$: the perceptual distance is defined by the *quantity of motion*. Remember that this distance is, until other constraints are used, defined up to scaling factors along the different axes.

Then the algorithm fills the distance matrix using Floyd’s algorithm; and goes on as **MDS** does. We will not develop these points further.

6 Results

The experiments were run with increasing complexity: 1° no rotation and no distortion of the input, 2° distortion is added 3° rotation and distortion. Note that this means two different learning phases (flawless camera for 1° , distorted camera for $2^\circ/3^\circ$) and two conditions on the chosen data points (obtained by translation only $1^\circ/2^\circ$, using rotation 3°).

First, we test the algorithm on a small set of patches obtained with a fixed angular position of the camera, and positions corresponding to the nodes of a square net - i. e., they are obtained one from another through one-pixel translations. Results are shown in Fig. 4: a clearly meaningful embedding is found. The points of the embedding follow a square grid whose axes represent the quasi-orthonormal basis.

Then we apply a distortion filter on the camera. We use a pixel-by-pixel multiplication with a “Gaussian \times sinus” function to produce the data. The period of the sinus function is such that two quadrants of the screen are lightened while two are darkened. Our algorithm is not affected, since

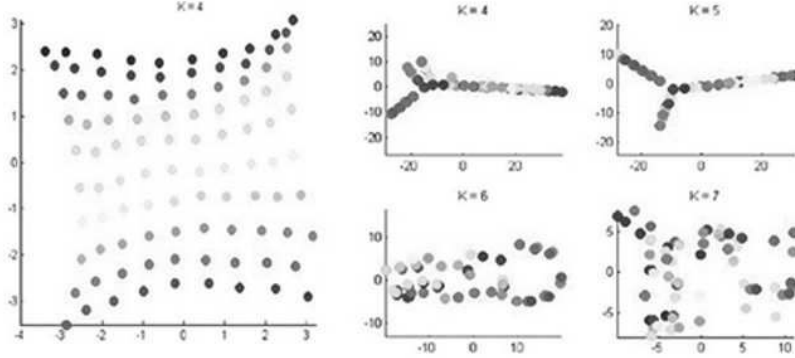


Figure 5: Again, the camera producing the data does not rotate, and the grid of patches is similar to the one in Fig. 4. *Left*, our algorithm is not affected and outputs a neat embedding. *Right*, Isomap, tested with different values of the neighborhood parameter (number of neighbours=4,5,6,7..), is affected by the distortion introduced in the image.

it learns the corresponding predictors all the same. As an indication, we compare in Fig. 5 the results with **Isomap**’s. Unsurprisingly, distortion of the image changes the Euclidean metric in the pixel space.

We go now for the full test. Patches are taken from the world with various positions *and* orientations of the camera. After learning, the reduction module outputs a 3-dimensional embedding - i.e. the first three eigenvectors obtained by the MDS are used. Results are shown in Fig. 6. The relevant structure of the 3D grid, “translation times translation times rotation”, is clearly exhibited. Patches obtained from the same angle are mapped onto horizontal planes, and these are organized along square grids; when **Isomap** utterly fails.

7 Discussion

7.1 Efficiency

The algorithm was not designed for efficiency. Considering that for 21×21 pixels patches the computation was already time-consuming, we admit it is, as such, limited. The core algorithm’s computational cost increases in $O(n^2 \cdot P^2)$, n the number of points, P the size of the patch side: every pair of points has to be matched with a combination of predictors. Above all, the memory load for the learning quickly becomes huge with N (initial number of actions) or P ($\sim N^2 \cdot P^4$).

Yet this does not mean that the path is a dead-end. First, the heaviest costs are attached to the learning phase, which occurs just once. Second, we successfully conducted some trials with bigger patches (31×31 pixels)

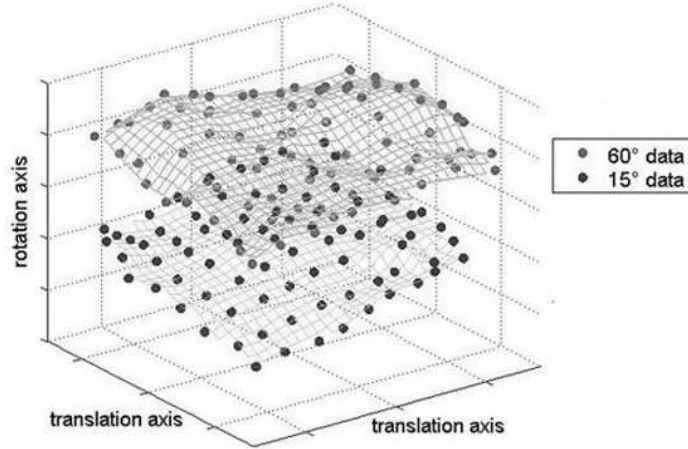


Figure 6: Data is obtained by the camera with a variation of the three parameters (x, y, θ) . For clarity, we represent only a subset of this data: points whose initial coordinates in the 2D world are of type (i, j, α) where i and j are integers and $\alpha = 15^\circ$ or 60° . The resulting embedding shows as expected the appearance of a "millefeuilles", where the vertical axis perfectly relates to the rotation angle and the horizontal axes to x and y .

by constraining the predictors to be a banded or sparse matrix. Indeed, with reasonable assumptions, local information seems enough to predict the transformation of a given pixel when undergoing a small motion. Sparse matrices speed up processes and spare memory to a large extent.

As far as action space is concerned, one solution could be the implementation of an incremental search. The algorithm would start with a small number of actions that would evolve. Successive stages of competence are widely used in developmental research, and can relate to the Piagetian sensorimotor stages that infants allegedly undergo during development (see Piaget [11]).

Other steps of the program are distinctly improvable. For instance, the similarity matrix is filled with Floyd's algorithm for basic coding convenience. In this paper we simply intended to propose a new look upon a well-known problem, putting forward an "active" notion of distance. Efficient, classically known methods can then take over from the similarity matrix we compute.

Note that while the numerical implementation of our idea raised various problems, it is unclear whether the difficulties would be the same in the case of a biological organism, or even an artificial but analogical agent.

7.2 Motivation

Aside from practical interest, our embodied view on DR takes inspiration from the field of human perception. There is a long history of theories

about the sensorimotor grounding of perceptual experience, reaching back to Helmholtz [9] or Poincaré; later Gibson [10]’s ecological theory in psychology and Piaget [11]’s account of infant development built on the crucial role played by active behavior in perception. Philosopher Merleau- Ponty [12] and followers (see for instance Dreyfus [13], Crossley [14]) emphasized the importance of one’s moving body to “measure the outside”. More recently, in the field of psychophysics, O’ Regan and Noë. [15, 16] push the approach further and describe what they call “sensorimotor contingencies”, laws ruling the interaction between an organism and the world: perception is described as cognitive access to the mastery of these. In this view our experience of the world is to be explained by studying the laws governing our interaction with it, and *not* by the way the data are encoded by the sensory apparatus.

7.3 Conclusion: Active Understanding of the World

The type of distance we have introduced constitutes one choice, partly inspired by work on human perception (Sec.7.2), but other possibilities exist. Information theory or statistical considerations on the input space are widely used to define or improve metrics relevant to the problem (see for instance [23], [24]). Other approaches assume that data come with ‘relevance information’ and build a consequently biased notion of distance([25]).

Our belief is that a metric such as ours is particularly useful in the case of an embodied artificial agent, since it leads to build a model of the world as experienced by the agent’s acting self. We have illustrated this approach for the very particular case where the actions involved are spatial displacements such as those encountered in the domain of 3D-scene reconstruction from sample pictures.

It would be interesting to try to extend this approach to the case when actions involved are not spatial displacements but have some other combinatorial structure. For example it might be possible to exploit the relationships between the transformations induced by blurring and zooming. In the domain of phoneme recognition, it might be possible to use constraints about the way the phonatory apparatus produces sounds.

Acknowledgment

This work was supported by the EC under contract number FP6-004250-CoSy.

Download

The code for the program can be downloaded at <http://nivea.psychology.univ-paris5.fr/~glanois/>

References

- [1] J-B. Tenenbaum, V. Silva and J-C. Langford (2000): A Global geometric Framework for Nonlinear Dimensionality Reduction. *Science*,

290(5500):2319-2323.

- [2] S.T. Roweis and L.K. Saul (2000): Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500):2323-2326.
- [3] M.Belkin, P.Niyogi (2003): Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Computation*, vol. 15, pp. 1373-1396.
- [4] P. Demartines and J. Herault (1997): Curvilinear Component Analysis: A Self-Organizing Neural Network for Nonlinear Mapping of Data Sets. *IEEE Trans. Neural Networks*, vol. 8, pp. 1197-1206.
- [5] B. Schölkopf, A. Smola and K.-R. Müller (1999): Kernel Principal Component Analysis. In B. Schölkopf, C. J. C. Burges, & A. J. Smola, editors, *Advances in kernel methods—Support vector learning*, pp. 327–352. : MIT Press, Cambridge, MA.
- [6] M. Brand (2003): Charting a manifold. *Advances in Neural Information Processing Systems 15*. MIT Press, Cambridge, MA.
- [7] Y. Bengio *and al.* (2004): Spectral Dimensionality Reduction. *CIRANO Working Papers 2004s-27*, CIRANO.
- [8] T.Kohonen (1988) Self-Organization and Associative Memory. *Springer-Verlag, Berlin*.
- [9] H. Helmholtz (1878): The Facts of Perception. *Selected Writings of Hermann Helmholtz*, Wesleyan University Press
- [10] J.J.Gibson (1979): The Ecological Approach to Perception. *Houghton Mifflin, Boston*
- [11] J. Piaget (1937) La construction du réel chez l'enfant. *Delachaux and Niestle, Paris*
- [12] M. Merleau- Ponty (1945): Phénoménologie de la perception. *Gallimard*
- [13] H.L.Dreyfus (2002): A Phenomenology of Skill Acquisition as the basis for a Merleau-Pontian Non-representationalist Cognitive Science. In *Proceedings of the Int. Conf. Foundations and the Ontological Quest.*, Rome, Vatican City
- [14] N.Crossley (2001): The phenomenological habitus and its construction. In *Theory and Society*, Volume 30, Number 1, pp. 81-120.
- [15] J. K. O' Regan and A. Noë (2001): A Sensorimotor Account of Vision and Visual Consciousness. *Behavioral and Brain Sciences*, 24(5), 939-1011.

- [16] A. Noë (2004): Action in Perception. *The MIT Press*
- [17] D. Philipona *and al.* (2004): Perception of the structure of the physical world using unknown sensors and effectors. In Kearns, M., Solla, S., Cohn, D., editors, *Advances in Neural Information Processing Systems 15*. MIT Press, Cambridge, MA.
- [18] R. P. N. Rao and D . L. Ruderman (1999): Learning Lie Groups for Invariant Visual Perception. In Kearns, M., Solla, S., Cohn, D., editors, *Advances in Neural Information Processing Systems 11*, pages 810-816. MIT Press, Cambridge, MA.
- [19] Y. Bengio and M. Monperrus (2005): Non-Local Manifold Tangent Learning. In Saul, L., Weiss, Y., and Bottou, L., editors, *Advances in Neural Information Processing Systems 17*. MIT Press, Cambridge, MA.
- [20] R. Pless (2004): Differential structure in non-Linear Image Embedding Function. In *IEEE CVPR Workshop: Articulated and Nonrigid Motions*, Baltimore.
- [21] R. Souvenir and R. Pless (2005): Isomap and Nonparametric Models of Image Deformation. In *IEEE Workshop in Motion and Video Computing*, Breckenridge, CO.
- [22] P. Y. Simard *and al.* (2001) Transformation Invariance in Pattern Recognition - Tangent Distance and Tangent Propagation. *International Journal of Imaging System and Technology*, Volume 11, Issue 3, pp. 181-194
- [23] H. Snoussi and A. Mohammad-Djafari (2002): Information geometry and prior selection. In C. Williams, Ed., *Bayesian Inference and Maximum Entropy Methods*, pp. 307–327, MaxEnt Workshops, Amer. Inst. Physics.
- [24] E. Xing *and al.* (2002): Distance metric learning with application to clustering with side-information. In *Advances in Neural Information Processing Systems*, MIT Press.
- [25] J. Peltonen, A. Klami and S. Kaski (2004): Improved Learning of Riemannian Metrics for Exploratory Data Analysis. *Neural Networks*, vol. 17, pp. 1087-1100

Similarity judgments and sensorimotor interactions

Philipona, D. and Glanois, F. and O'Regan, J.K.

1 Introduction

We naturally express perceptual similarity judgments between stimuli, both categorical and quantitative. We do so about colors, shapes, faces, pitches, timbers, phonemes, and so on. These judgments are not related with our raw sensory inputs in any simple way, and they often result from a complex learning process taking place over development. Many psychophysical researches have been devoted to identify and analyse similarity judgments in humans, and it is widely acknowledged that they show some level of consistency both within and across subjects. There is a long standing debate, though, about their origin.

It has been suggested by different authors that the origin of perceptual judgments lies in the structure of sensorimotor interactions of the subject with its environment. One advantage of such theories is to acknowledge that perceptual judgments inherently depend *both* on the environment properties and on the physiological constraints of the organism. Here we show the example of a system implemented to generate similarity judgments about its sensory inputs, based on its interactions with the environment.

The general approach is three steps. First, the system interacts randomly with its environment and grows a finite repertoire of elementary actions having some simple combinatorial structure. Second, the system develops classifiers to recognize sensory evolution resulting from these actions. And third, a local notion of similarity between sensory inputs is constructed based on these classifiers, and extended to a global notion of similarity.

The extension of the local notion of similarity to a global one is closely connected to so called nonlinear dimensionality reduction techniques. Classically, however, reduction techniques such as Isomap rely on a local notion of similarity grounded in the encoding format of sensory inputs. This is what we call a ‘passive’ viewpoint. In the present work on the opposite, the crux of the approach is to derive the local notion of similarity from the experimentation of the system used to acquire the data.

The paper is organized as follow. In section 2, we review the background issues. In section 3, we detail an implementation of our approach for a simple system and environment. In section 4, we present the results of simulations.

In conclusion, we discuss related works, the generality of the approach, and the question of (low-dimensional) representations.

2 Background issues

2.1 Perceptual similarity and dimensionality reduction

What is perceptual similarity? To introduce the concept, one may refer to different tasks set up by psychologists: verbal similarity ratings, spontaneous analogy making, cross-cultural linguistic categorization, etc. Data collected from such tasks are referred to as “data about perceptual similarity”, and the tasks can involve elementary stimuli such as color, sounds, or compound stimuli such as faces, shapes, pictures, speech, musics (see for instance [25, 22] for particular examples, or [39, 44] for somewhat general viewpoints on the issue of similarity). These data often show some level of consistency both within and across subjects, and not less often their explanation is a subject of debate between specialists. As an example, there exist different proposals for the origin of the perceptual categorization of colors: specificities of neuronal representations [7], ecological constraints [34], linguistic and social interactions [9]. While the proposals may be compatible, it is not clear how to coordinate them, and actually they are commonly advertized as autonomous theories.

During the last decade, some researchers have suggested that “the perceptually meaningful structure” of sensory data lied within the manifold structure of a set to which these data belongs [13, 41, 24, 42, 35]. For instance, tasks based on perceptual similarity between pictures of faces – interpolation, extrapolation, analogy making – make sense, it has been argued, relative to the manifold structure of a set of pictures of faces (Figure 1.a). It was actually argued that these tasks should be made relative to *geodesics* of the manifold, i.e. the shortest paths within the set of pictures. The notion of geodesics, however, rely on the exploitation of a riemannian metric on the manifold of observations, a metric which is typically unwarranted.

Nonetheless, the suggestion drawn inspiring connections between researches on perceptual similarities and a standard issue of the computational learning field: the search of low-dimensional structure in sets of high-dimensional observations. In computational research, the issue is usually formalized as the construction of an embedding of observations within a low-dimensional representational space, hence the name of ‘dimensionality reduction’. A wealth of learning algorithms have been developed to address this goal [10, 35, 42, 3, 11, 5, 8], and Isometric feature mapping (Isomap) is arguably the simplest of them all. First, a graph of neighborhood relationships between points is constructed to reflect the topology of the dataset. Weights are then assigned to each link of the graph so as to indicate lo-

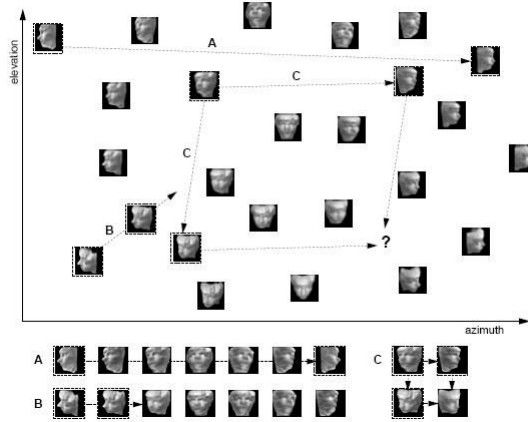


Figure 1: Interpolation, extrapolation and analogy making from dimensionality reduction (sketch from [43])

cal distances between neighboring points, as computed from the Euclidean norm in the observation space. From this weighted graph, a similarity matrix M is then constructed such that M_{ij} equals the weight of the shortest path within the graph between points i and j . And finally, multidimensional scaling is used to build a low-dimensional representation of the observation dataset from the similarity matrix M .

2.2 Perceptual similarity and sensorimotor theories

Reduction techniques such as Isomap critically depend on the local notion of similarity¹ they consider between observations. But what could be a “perceptually meaningful” notion of local similarity? Consider, for instance, a picture of alternating black/white vertical stripes; compare it with the same picture translated horizontally by a few pixels. How similar are they? If we apply the Euclidean norm, which involves a pixel by pixel computation, the dissimilarity is quite significant. Yet, for an embodied system, it makes sense to consider them as very similar: if an active system has to judge whether two stimuli are almost identical, its judgments have to take into account the uncertainty resulting from sensor noise, but also, just as importantly, the uncertainty on sensory data resulting from its own actions.

There is another concern about merely relying on the encoding format of sensory inputs to determine similarity. Sensory systems show significant physiological variations between individuals, they further evolve as well as suffer damages during individuals’ life, and can be artificially altered in different ways. Yet similarity judgments show a greater consistency across

¹Throughout this letter, the terms ‘similarity’ and ‘distance’ will be used interchangeably, but it is well known that this identification is problematic, see for instance [38, 36].

perceivers than would be expected from the various neuronal encoding format of sensory inputs. There is good agreement that sensorimotor experimentation plays a role in this consistency, if only because this experimentation drives the development of the central nervous system [20, 37]. But some researchers have pushed forward a stronger claim, namely that perceptual judgments are *about* the structure of sensorimotor interactions [27].

There is actually a long history of theories about the sensorimotor grounding of perceptual judgments, especially with respect to spatiality reports, reaching back to Helmholtz and Poincaré [45, 33]. Gibson’s ecological theory in psychology [16] and later Piaget’s account of infant development built on the crucial role played by active behavior in perception [30]. The idea seduced as well influential philosophers [23, 14], who emphasized the importance of one’s moving body to “measure the outside”. The approach we suggest in this paper is an attempt towards a computationally tractable formulation of such sensorimotor views about perceptual similarity.

2.3 Perceptual similarity and the structure of actions

If we consider a set of pictures of a scene seen from different viewpoints, and ask a subject to extrapolate a series of stimuli such that there always is the same viewpoint change between two successive stimuli, there is little doubt that the subject will actually point out the stimulus that is consistent with a new iteration of that *same viewpoint change*. In other words, perceptual similarity will then refer to the fact that sensory data are consistent with the repetition of a given *action*, not to some measure of the almost equality of sensory data encodings. Defining the Riemannian metric required by dimensionality reduction techniques, in this case, is therefore achieved by selecting a finite repertoire of actions (Fig. 1.b).

Grounding perceptual similarity in sensorimotor interactions rather than sensory encoding provides a natural explanation for the categorization of different perceptual changes: actions have a combinatorial structure (sequences of actions are actions), and not all actions play the same role with respect to that structure. For instance, consider a set of pictures of a scene seen from different viewpoints and focus. While it is not clear from sensory data themselves why one should consider as perceptually special those changes that result from a pure change of viewpoint and those that result from a pure change of focus, the combinatorial decoupling of the corresponding actions obviously retains an essential part of the answer.

When the action set has the very strong structure of a mathematical group², it becomes clear, however, that one should not expect to find a procedure to isolate decoupled actions sets that is both general and simple:

²in addition of the fact that sequences of actions are actions, there is a null action and any action can be reverted

groups can have notoriously complex structures, and general methods of decompositions are accordingly intricate. In the following, we suggest a rather crude process of decomposition that detects actions subsets that form basic subgroups of actions having particular properties with respect to commutativity: the so called *center* and *derived subgroup* [18]. It turns out that, in the previous example, these subgroups correspond to pure focus changes and pure translations of the viewpoint, respectively. They provide thus the decomposition of the actions set that seems to us very natural.

We will not elaborate about possible motivations for an organism to grow a repertoire of actions having good combinatorial properties, yet we note that symmetries have a big impact on the efficiency of planning [15], and that commutativity is known to make planning easier because it allows to prune a considerable number of action sequences from the search space of the planner [19]. It seems therefore that an organism able to grow such a repertoire of actions will get some advantage over an organism that is not.

3 Sensorimotor reduction of dimensionality

Here we detail our approach on a specific example. In section 3.1, we describe the sensorimotor interactions that our system can engage in, and the basic learning ability we assume it has. In section 3.2, we then explain how this ability can be used to generate similarity judgements about the system's sensory inputs.

3.1 The simulated system

3.1.1 Sensorimotor interactions

The sensorimotor interactions we consider here are the ones that would be faced by a system constituted of a pinhole camera moving in a flat two-dimensional space (Fig. 2). The camera can move around and zoom in or out: the state of the system can be described by real parameters (x, y, θ, f) , where (x, y) is the position of the camera in the space, θ is its orientation, f is the focal length of the camera, and the multidimensional action $a = (a_1, a_2, a_3, a_4)$ is the following evolution of the system:

$$\begin{cases} x_{t+1} &= x_t + a_1 \cos \theta_t + a_2 \sin \theta_t \\ y_{t+1} &= y_t - a_1 \sin \theta_t + a_2 \cos \theta_t \\ \theta_{t+1} &= \theta_t + a_3 \\ f_{t+1} &= f_t + a_4 \end{cases} \quad (1)$$

We assume that actions are determined by multidimensional motor commands $m = (m_1, \dots, m_x) \in \mathbb{R}^x$, i.e. $a = \varphi(m)$ where φ is some smooth function.

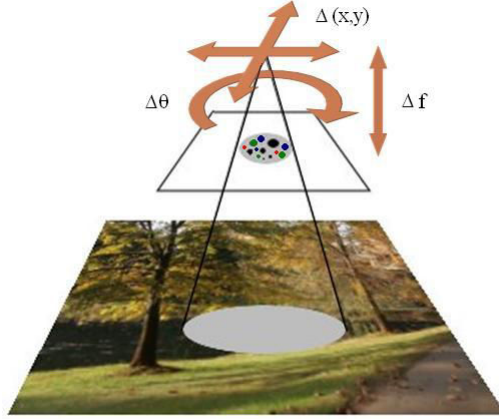


Figure 2: a) The simulated sensorimotor system constituted of a pinhole camera moving in a flat two-dimensional space. The camera can move around and zoom in or out: the state of the system can be described by real parameters (x, y, θ, f) . The artificial retina is constituted by photoreceptor with different sensitivities, both in amplitude and spectrum (represented by the dots of different sizes and color), randomly distributed.

The snapshot taken by the camera is then determined from the location of the camera over a large gray-scale image, and the focal length. To evaluate the influence of the encoding format of sensory data on the similarity structure generated by the system, the snapshot will be sampled in a spatially non-uniform way, with photoreceptors having different sensitivities to the light level. Herein, we note $s = (s_1, \dots, s_x) \in \mathbb{R}^x$ the encoded sensory data.

This example is rather naïve, yet it already exhibits several relevant difficulties. First, the sensory inputs are high-dimensional (40×40 pixels, i.e. 1600-dimensional vectors), imposing computational constraints. Second, the action group is the product of two decoupled action sets: that of spatial movements and that of focus changes. And third, while the combinatorial structure of movements in a 2D space is simpler than that of movements in a 3D space, the essential features of a non trivial subgroup of actions are already present there, so that the two decoupled action sets are not both trivial (the combinatorial structure of focus changes, of course, *is* trivial).

3.1.2 Basic learning ability

Given a sequence of motor commands $[m(1) \dots m(k)]$, we will say that the motor command m cancels the sequence if, for most states of the system, the sensory inputs after issuing the commands sequence $[m(1) \dots m(k).m]$ are the same as before the commands sequence. The fundamental learn-

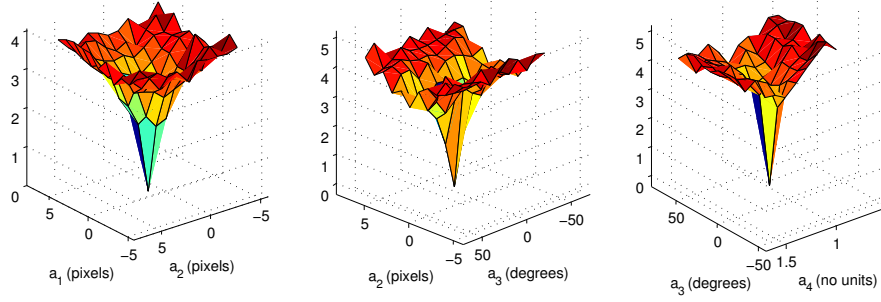


Figure 3: Adjustment of canceling commands: average sensory change as a function of the actions coordinates (a_1, a_2) , (a_2, a_3) and (a_3, a_4) (see Eq. 1 for the meaning of these coordinates).

ing ability of the system will be to adjust an estimate of such cancelation commands.

In the current simulation, this adjustment is realized by the active experimentation of the neighboring commands of the estimate – herein we will say *variations* –, produced by adding a gaussian noise to the command. Let \tilde{m} be a random variation of the starting estimate \hat{m} of m . If $[m(1) \dots m(k). \tilde{m}]$ decreases the average sensory change as compared to $[m(1) \dots m(k). \hat{m}]$, then \tilde{m} becomes the new estimate of m . If it increases the average sensory change, it may also become the new estimate but with a low probability, that lowers with time. The process repeats until the average sensory change is below some threshold. In other words, we consider a simulated annealing process [21] minimizing in m

$$E([m(1) \dots m(k).m]) = \frac{1}{N} \sum_{i=1}^N \|\Delta s(i)\|^2 \quad (2)$$

where $\{\Delta s(i), i = 1 \dots N\}$ is a set of sensory changes produced from the commands sequence $[m(1) \dots m(k).m]$ issued in N different states of the system.

Fig. 3 shows the profile of E as a function of the action coordinates associated with the commands sequence. It is clear from it that this annealing process will work efficiently as long as $[m(1) \dots m(k). \hat{m}]$ is a small action, i.e. as long as the starting estimate of m is not too far away from a command that cancels the sequence, hence the term of “adjustement”.

As part of the adjustment ability of the system, we will assume that the system can rely from the start on a reference resting command m^* that produce little change of sensory inputs.

3.2 Generating similarity judgments

Our approach for generating a structure of similarity judgments consists in three independent steps, that we summarize here and then detail in the rest of the section.

1. the system interacts with its environment and grows a finite repertoire of elementary actions chosen with respect to their advantageous combinatorial properties
2. the system develops classifiers that will allow it to recognize sensory evolutions consistent with actions of this repertoire
3. a local notion of similarity between sensory inputs is defined from the classifiers, then extended to a global notion of similarity.

To display the global similarity structure thus obtained, we then use multi-dimensional scaling.

The main novelty of this letter is to argue for grounding step 3 upon steps 1 and 2, and to provide an example of how the three steps can be implemented so as to yield altogether coherent results. However, there is considerable freedom in the way each step could in general be implemented.

3.2.1 Step 1: Growing the action repertoires

To develop its action repertoires, the system will grow actions by bootstrapping its adjustment ability. A pair (m, m') of commands that cancel each others and are not too far away from m^* can be grown into a serie of larger and larger actions by successive adjustments: start with $m(1) = m$ and $m'(1) = m'$, then defines iteratively $m(k+1)$ as the adjustment of $m(k)$ that cancels $[m'(k).m']$, and $m'(k+1)$ as the adjustment of $m'(k)$ that cancels $[m(k+1)]$. In such a serie, we will say that $m(k)$ is the *power* k of m , and k is the *exponent* of $m(k)$.

The smoothness of φ implies that variations of the reference command m^* can be cancelled by commands that are themselves variations of m^* , therefore the system can generate pairs of commands (m, m') that cancel each others: let m be a random variation of m^* , and m' be the adjustment of m^* that cancels $[m]$. By producing in this way n random pairs of commands that cancel each others, the system can produce a first finite repertoire of commands $\mathcal{R} = \{m(i), i = 1 \dots n\}$, with an association table for the commands that cancel them, and another for their powers up to some order.

As a matter of normalization, the system retains n commands that induce approximately the same, predefined, average sensory change as measured by $E(m)$ – herein we will say *normalized* motor commands. Because of this choice, the Euclidean norm on sensory encodings is going to affect somewhat our results (though little, see Sec. 4). The structure of actions should in

theory provide constraints for a more principled choice, but it seems relevant to investigate first such a simpler approach.

From this basic repertoire of actions that the system has learnt to cancel and increase, we will now consider the issue of deriving from it two other repertoires adapted with respect to commutativity properties of actions.

Let us state briefly two definitions: the center of a group (G, \cdot, e) is the set of $a \in G$ such that $a \cdot b \cdot a^{-1} \cdot b^{-1} = e$ for all $b \in G$, the derived subgroup is the group generated by all elements of the form $a \cdot b \cdot a^{-1} \cdot b^{-1}$, where $a, b \in G$. These two groups are key elements of standard decompositions of groups with respect to their combinatorial structure.

To generate a normalized action in the center of the action group, the system once more relies on simulated annealing to find a command m minimizing

$$E_C(m) = \max_{m(i) \in \mathcal{R}} E([m(i).m.m'(i).m'])$$

where m and m' , $m(i)$ and $m'(i)$, cancel each others. The detailed procedure is the following. Starts with generating a normalized variation m of m^* , find a cancelling command m' of $[m]$ by adjusting m^* , and evaluate $E_C(m)$. Then iterate: generate a normalized variation \tilde{m} of m , find a canceling command \tilde{m}' of $[\tilde{m}]$ by adjusting m' , evaluate $E_C(\tilde{m})$, and always retains \tilde{m} as a current estimate of an element of the center if $E_C(\tilde{m}) < E_C(m)$, or retains it with a low probability that decrease with time in the other case.

To generate a normalized action in the derived subgroup of the actions combinatorial structure, the system uses the following procedure: let m be a command adjusted from m^* so as to cancel the command sequence $[m(i).m(j).m'(i).m'(j)]$ for randomly chosen $m(i), m(j) \in \mathcal{R}$, and let m' be adjusted from m^* so as to cancel $[m]$. If the average sensory change for the resulting command is below the normalization threshold, then start again the same procedure but using the powers of $m(i)$ and $m(j)$ instead of $m(i)$ and $m(j)$.

From the ability to generate normalized actions in the center and derived subgroups of actions, the system can create two additional repertoires, \mathcal{R}_{Center} and $\mathcal{R}_{Derived}$, with association tables for the cancelling commands and powers up to some order.

3.2.2 Step 2: Learning classifiers

Each command m of a repertoire is associated with specific statistics of sensory evolution: the statistics of the set $\{(s(t), s'(t)), t = 1 \dots N\}$, where the $s(t)$ are sensory inputs for N different states of the system, and the $s'(t)$ are the sensory inputs after issuing m in these states. From these, the system can learn classifiers to recognize sensory evolution that are consistent with the action. We consider in the current example a case of gaussian classifiers, and therefore we briefly recall what gaussian classifiers are.

Figure 4: Similarity structure with respect to the derived subgroups of actions (translation) and the center of actions (focus change).

Figure 5: Influence of sensory data encoding: comparison with Isomap.

Let $\{m(i), i = 1 \dots n\}$ be a set of commands and $\mu(i)$, $\Sigma(i)$ be the means and covariance matrices of the sensory evolutions associated with each command. The gaussian likelihood that $x = (s, s')$ corresponds to a sensory evolution resulting of action $m(i)$ is:

$$p_i(x) \propto |\Sigma(i)|^{-1} \exp \left[(x - \mu(i))^t \Sigma(i)^{-1} (x - \mu(i)) \right]$$

and the most reasonable choice for i is the one such that $p_i(x)$ is maximal over $i = 1 \dots n$. Gaussian classifiers are well known to be implementable by neural networks, but in our simulation we merely computes likelihoods from the previous formula.

Instead of using a global classifier that would exploit all covariations between all the coordinates of sensory evolutions and therefore require the inversion of a huge matrix, we will consider a large collection of local classifiers, each voting for a particular action. If there is an action that receives a large proportion of the votes, then it is considered to be compatible with the sensory evolution. If votes are much dispersed, then the sensory evolution is considered not to be compatible with any action.

3.2.3 Step 3: Local and global similarities

Given a set of sensory data, the system can now exploit a “perceptually meaningful” notion of local similarities so as to generate a global structure of similarities.

Using the classifiers associated with its different action repertoires, the system can establish different graphs of neighboring relationships: s and s' are “perceptual” neighbors for a given repertoire if the sensory evolution (s, s') is consistant with a command of this repertoire or one of its power, and each edge of the graph is assigned as a weight the corresponding exponent. A global notion of similarity is then derived from this graph in the same way as Isomap does, using Floyd’s shortest path algorithm.

4 Results

5 Conclusion

5.1 Related works

The problem of defining measures of perceptual similarity is well known and subject of much work by computer scientists: it is indeed the key ingredient for retrieval routines in multimedia databases. All these works, however, are not concerned with how a sensorimotor agent could build a notion of perceptual similarity: they rather look for a program that outputs from scratch and at low computational cost the notion of similarity that we humans report.

In the specific case of dimensionality reduction, it is not new to stress that methods based on unwarranted local distances produce parameters that “do not have explicit meaning”, and that the choice of these distances therefore deserve consideration [32, 40]. But there again, not much work have been devoted to the question of building a notion of local distances from the sensorimotor interactions of an agent.

Action Respecting Embedding (ARE) [6], an algorithm that constraint dimensionality reduction with respect to action labels given between sensory data, is close in spirit with the current approach. ARE however, is not concerned with the determination of an action repertoire, nor with the issue of *inferring* action labels from sensory data pairs. In fact, it seems that ARE could provide a relevant alternative for our simple implementation of the third step of our approach.

Our own previous work [28, 29] discusses how an organism can interpret the spatiality of its environment from the structure of its sensorimotor interactions, but it does not deal neither with any inference abilities. Works about learning the changes of sensory signals from uninterpreted captors that result from various motor commands [31, 26], or about general transformations for sensory data [17, 4], could provide more powerful implementations for the classifiers of the second step. At last, the first step would presumably benefit from researches about algorithms decomposing so called “black-box” groups, i.e. groups specified by a “black-box” performing on command some group operations [1, 2].

5.2 Generality of the approach

It is not clear how general is the approach we suggest in the letter: the notion of action that it builds upon here is obviously of a spatial nature. What would be, for instance, the combinatorial structure of actions for sounds, in particular for speech production? At best, it seems that in such case the assumption of a group structure for actions would have to be weakened.

Quite surprisingly, though, our approach suggests interesting ideas for the case of colors. The so called “perceptual dimensions” of colors – value, saturation and hue – are essentially dimension of sensory changes that are consistant with actions having a very different role in the combinatorial structure of all actions: change in value is consistant with a change in the aperture of the diaphragm, change in saturation is consistant with a change in the relative orientation of the eye with respect to a shiny surface, and change in hue is consistant when manipulating colored surfaces under different illuminants. Such a naïve account of perceptual dimensions, it seems, would fall short of explaining several well know effects related to these – e.g. the Burns effect or the Bezöld-Bruck effect. But of course, we can naturally expect that the classifiers inferring actions from sensory evolutions can go wrong in some cases, or even exhibit a systematic bias, without this undermining the whole story.

What should be counted in favor of sensorimotor approaches is that they give an intelligibility to the very notion of what could *be* a perceptual dimension, without appealing at some point to a yet to be discovered mysterious “psycho-physiological process generating sensations of different qualities”.

5.3 The question of representations

It is often argued by some psychologists that the structure of perceptual similarity judgments in humans follows straightforwardly from the existence of things called *representations* [12]. Dimensionality reduction, introduced in computer science as the problem of finding descriptions of a dataset in terms of a few parameters, is often taken to provide a concrete example of what such putative representations are and how they could be derived [13].

While we do not intend to discuss the argument that low-dimensional neuronal encoding may facilitate learning because of the curse of dimensionality, nor the claim that specific representations may be important for acting out certain cognitive abilities, the claim that the structure of similarity judgments unveils the existence of such representations do seem to us farfetched. In our simulation, there is a system faced with sensory data whose intrinsic dimensionality is low, a system whose functioning is tuned to exploit this fact so as to express similarity judgments (hence, potentially, to fullfil all perceptual tasks related to similarity judgments), but there is no low-dimensional representations in the “brain” of our system. The only low-dimensional representations in this work are the ones derived in Fig. x and y in the address of the reader, so that he can get a handle on the structure of similarity judgments using a spatial analogy.

References

- [1] L Babai and R. Beals. A polynomial-time theory of black box groups. In *Groups: St Andrews 1997 in Bath*, 1999.
- [2] L Babai, WM Kantor, PP Pálffy, and A Seress. Black-box recognition of finite simple groups of lie type by statistics of element orders. *Journal of Group Theory*, 2002.
- [3] Mikhail Belkin and Partha Niyogi. Laplacian Eigenmaps for Dimensionality Reduction and Data Representation. *Neural Comp.*, 15(6):1373–1396, 2003.
- [4] Y. Bengio and M. Monperrus. Non-local manifold tangent learning. In Weiss-Y. Saul, L. and L Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2005.
- [5] Yoshua Bengio, Olivier Delalleau, Nicolas Le Roux, Jean-François Paiement, Pascal Vincent, and Marie Ouimet. Learning eigenfunctions links spectral embedding and kernel pca. *Neural Comput*, 16(10):2197–2219, Oct 2004.
- [6] Michael Bowling, Ali Ghodsi, and Dana Wilkinson. Action respecting embedding. In *ICML '05: Proceedings of the 22nd international conference on Machine learning*, pages 65–72, New York, NY, USA, 2005. ACM.
- [7] E.J. Chichilisky and Brian A. Wandell. Trichromatic opponent color classification. *Vision Research*, 39(20):3444–58, 1999.
- [8] Denis Chigirev and William Bialek. Optimal manifold representation of data: an information theoretic approach. *Advances in Neural Information Processing Systems*, 16, 2004.
- [9] J. Davidoff. Language and perceptual categorisation. *Trends Cogn Sci*, 5(9):382–387, Sep 2001.
- [10] Pierre Demartines and Jeanny Hérault. Cca: Curvilinear component analysis. In *15th workshop GRETSI*, 1995.
- [11] David L Donoho and Carrie Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proc Natl Acad Sci U S A*, 100(10):5591–5596, May 2003.
- [12] Shimon Edelman and Sharon Duvdevani-Bar. Similarity, connectionism, and the problem of representation in vision. *Neural Computation*, 9(4):701–721, 1997.

- [13] Shimon Edelman and Nathan Intrator. Learning as extraction of low-dimensional representations. In D. Medin, R. Goldstone, and P. Schyns, editors, *Mechanisms of Perceptual Learning*. Academic Press, november 1997.
- [14] Gareth Evans. *The varieties of reference*. Oxford University Press, 1991.
- [15] Maria Fox and Derek Long. The detection and exploitation of symmetry in planning problems. In *IJCAI*, pages 956–961, 1999.
- [16] James J. Gibson. *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, 1979.
- [17] David B. Grimes and Rajesh P.N. Rao. Bilinear sparse coding for invariant vision. *Neural Computation*, 17:47–73, 2005.
- [18] Morton Hamermesh. *Group Theory and Its Application to Physical Problems*. Courier Dover Publications, 1989.
- [19] Patrik Haslum and Hector Geffner. Admissible heuristics for optimal planning. In *Artificial Intelligence Planning Systems*, pages 140–149, 2000.
- [20] R. Held and A. Hein. Movement-produced stimulation in the development of visually guided behavior. *J Comp Physiol Psychol*, 56:872–876, Oct 1963.
- [21] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, May 1983.
- [22] S. McAdams, S. Winsberg, S. Donnadieu, G. De Soete, and J. Krimphoff. Perceptual scaling of synthesized musical timbres: common dimensions, specificities, and latent subject classes. *Psychol Res*, 58(3):177–192, 1995.
- [23] Maurice Merleau-Ponty. *Phénoménologie de la Perception*. Gallimard, 1945.
- [24] Baback Moghaddam. Principal manifolds and bayesian subspaces for visual recognition. In *ICCV (2)*, pages 1131–1136, 1999.
- [25] R. M. Nosofsky. Tests of an exemplar model for relating perceptual classification and recognition memory. *J Exp Psychol Hum Percept Perform*, 17(1):3–27, Feb 1991.
- [26] L. Olsson, C.L. Nehaniv, and D. Polani. From unknown sensors and actuators to actions grounded in sensorimotor perceptions. *Connection Science*, 18:121–144, 2006.

- [27] J. Kevin O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 2001.
- [28] David Philipona, J. Kevin O'Regan, and Jean-Pierre Nadal. Is there something out there ? Inferring space from sensorimotor dependencies. *Neural Computation*, 15(9), 2003.
- [29] David Philipona, J. Kevin O'Regan, Jean-Pierre Nadal, and Olivier J.-M.D. Coenen. Perception of the structure of the physical world using unknown sensors and effectors. *Advances in Neural Information Processing Systems*, 15, 2004.
- [30] Jean Piaget. *La construction du réel chez l'enfant*. Neuchâtel; Paris, 1937.
- [31] David Pierce and Benjamin Kuipers. Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92:169–229, 1997.
- [32] Robert Pless. Differential structure in non-linear image embedding functions. In *CVPRW '04: Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04) Volume 1*, page 10, Washington, DC, USA, 2004. IEEE Computer Society.
- [33] Henri Poincaré. On the foundations of geometry. *The Monist*, 9:1–43, 1898.
- [34] B. C. Regan, C. Julliot, B. Simmen, F. Viénot, P. Charles-Dominique, and J. D. Mollon. Fruits, foliage and the evolution of primate colour vision. *Philos Trans R Soc Lond B Biol Sci*, 356(1407):229–283, Mar 2001.
- [35] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, December 2000.
- [36] Simone Santini and Ramesh Jain. Similarity measures. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9):871–883, 1999.
- [37] J. Sharma, A. Angelucci, and M. Sur. Induction of visual orientation modules in auditory cortex. *Nature*, 404(6780):841–847, Apr 2000.
- [38] RN Shepard. Toward a universal law of generalization for psychological science. *Science*, 237(4820):1317–1323, 1987.
- [39] Roger N. Shepard. The perceptual organization of colors:an adaptation to regularities of the terrestrial world ? *Barlow/Cosmides/Toobu, The adapted Mind*, pages 495–532, 1992.

- [40] Patrice Simard, Yann LeCun, John S. Denker, and Bernard Victorri. Transformation invariance in pattern recognition-tangent distance and tangent propagation. In *Neural Networks: Tricks of the Trade, this book is an outgrowth of a 1996 NIPS workshop*, pages 239–27, London, UK, 1998. Springer-Verlag.
- [41] J. B. Tenenbaum. Mapping a manifold of perceptual observations. *Neural Information Processing Systems*, 10, 1998.
- [42] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.
- [43] Joshua B. Tenenbaum and William T. Freeman. Separating style and content with bilinear models. *Neural Computation*, 12:1247–1283, 2000.
- [44] Amos Tversky. *Preference, Belief, and Similarity* Selected Writings. MIT Press, 2004.
- [45] Hermann von Helmholtz. The origin and meaning of geometrical axioms. *Mind*, 1(3):301–321, 1876.

Troisième partie

Les couleurs de l'expérience

Chapitre 9

La structure perceptive des couleurs

9.1 Une approche sensorimotrice

Les propriétés spatiales et les couleurs, depuis au moins la distinction établie par John Locke entre qualités primaires et secondaires, occupent historiquement deux frontières du débat sur l'expérience sensorielle.

Dans son *Essay concerning human Understanding*, Locke développe une théorie des idées, “objets immédiats de la perception”, et du pouvoir des objets physiques de produire en nous ces idées, pouvoir qu’il appelle les *qualités* des objets. Les objets, dit-il, ont réellement une forme, une étendue, un nombre, un mouvement. Par contre, continue-t-il, les couleurs, les odeurs, les goûts, sont produits par l’action de particules sur notre système sensoriel, et il n’y a aucun rapport entre les propriétés physiques de ces particules, propriétés géométriques et numériques, et les idées qu’elles causent (Locke, 1690). Pour Locke, les propriétés spatiales sont une réalité physique, alors que les couleurs ne sont que des idées.

Peu de chercheurs, et certainement encore moins de philosophes, défendraient aujourd’hui dans le détail l’argumentation de Locke. Mais l’idée qu’une partie essentielle de nos affirmations sur l’espace trouvent leur origine dans la structure du monde physique, alors que celles sur les couleurs trouvent leur origine en nous, certainement dans notre système nerveux, reste largement répandue (Byrne and Hilbert, 1997). La dis-

inction entre qualités primaires et secondaire se trouve ainsi toujours enseignée dans certains manuels de neurosciences abordant le problème de la perception (e.g. [Kandel et al. \(2000\)](#)).

Il était donc intéressant d’aborder les deux questions de l’espace et des couleurs du point de vue d’une approche sensorimotrice de l’expérience. De ce point de vue, en effet, il n’y a pas de différence fondamentale entre les affirmations sur notre expérience de l’espace et des couleurs : elles trouvent leur origine à la fois dans les propriétés de l’environnement et de notre organisme, les dynamiques dans lesquelles nous pouvons nous trouver engagés étant déterminées à la fois par l’un et par l’autre.

Comme évoqué en section 5.2, l’idée que ce qu’on appelle l’espace est une contrainte existant sur la structure de nos capacités sensorimotrices n’est pas nouvelle. En fait, il s’agit même d’une idée naturelle : l’espace mesuré par un géomètre est sans ambiguïté une telle contrainte. A l’inverse, il n’existe semble-t-il aucune tentative d’expliquer la structure de l’espace des couleurs par une approche sensorimotrice. Les couleurs sont l’archétype de la sensation envisagée comme “imposée à nous”, indépendante d’une quelconque forme d’expérimentation, de savoir, ou d’interaction avec l’environnement. Les couleurs représentent l’énigme des sensations dans sa forme la plus directe.

Il a déjà été suggéré que les invariances des dépendances sensorimotrices permettent de proposer l’esquisse d’une caractérisation de l’obscurité (Sec. 6.5.4), peut-il en être de même pour les couleurs ? Ce qui suit propose une analyse basée sur le modèle dichromatique des propriétés de réflexion des surfaces. C’est un cadre qui peut laisser sceptique : la spécularité est généralement considérée comme un aspect secondaire dans la perception des couleurs. De fait, c’est avec des surfaces mates qu’il nous semble avoir les cas les plus élémentaires, les moins ambigus, de couleurs. Toutefois, la coïncidence qui existe globalement entre “la structure perceptive” des couleurs et la structure des dépendances indiquées par le modèle dichromatique interpelle, et, ne serait-ce qu’à ce titre, paraît mériter une analyse. De plus, la question de l’identification des propriétés des surfaces, pour laquelle la spécularité semble effectivement ne pas être exploitée ([Maloney, 2001](#)), est une question *a priori* différente de celle d’un jugement de similarité entre ces propriétés (mais voir également Sec. 9.2).

Selon le modèle dichromatique, la lumière réfléchie par une surface est la somme d’une lumière diffuse et d’une lumière spéculaire, dans des proportions qui dépendent de l’angle d’observation. Toujours selon ce modèle, le spectre de la lumière diffuse

est le produit du spectre de la lumière incidente avec une caractéristique de la surface nommée réflectance, et le spectre de la lumière spéculaire est le même que celui de la lumière incidente. Si l'on considère le cas d'un champ de lumière de direction donnée mais de composition spectrale variable selon la position, éclairant une surface de propriété chromatique uniforme, le modèle dichromatique s'écrit :

$$E_r(\lambda, p, \theta) = \alpha(\theta)E_i(\lambda, p) + \beta(\theta)E_i(\lambda, p)R(\lambda)$$

où $E_i(\lambda, p)$ la composition spectrale au point p de la lumière incidente, $E_r(\lambda, p, \theta)$ la composition spectrale au point p sous l'angle θ de la lumière réfléchie, et $R(\lambda)$ la réflectance de la surface (Fig. 9.1). Depuis [Lambert \(1760\)](#), il est généralement admis que $\alpha(\theta) = \cos(\theta)$ (où θ est l'angle avec la normale de la surface) est une approximation relativement réaliste pour un grand nombre de surface (mais voir cependant [Nayar and Oren, 1995](#); [Oren and Nayar, 1996](#)). Il existe par contre une variété de modèles dichromatiques, selon la forme considérée pour le facteur β : modèle de Phong, de Beckmann, de Heidrich - Seidel, de Cook - Torrance, pour n'en nommer que quelques uns. Ces différents modèles résultent de différentes hypothèses sur la distribution des microfacettes des surfaces. Il faut noter, enfin, que tous ces modèles sont un cas particulier de la notion de fonction de distribution de réflexion bidirectionnelle (BRDF) ([Koenderink et al., 1999](#); [Kelemen and Szirmay-Kalos, 2001](#)).

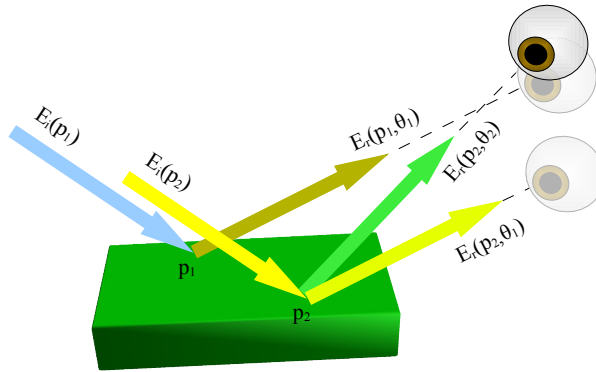


FIG. 9.1 – Modèle dichromatique. Exemple de dépendance entre la stimulation sensorielle et la position et l'orientation de l'oeil, dans le cas d'une surface uniforme, et d'un champ de lumière d'angle d'incidence donné mais de composition spectrale variable spatialement (variable λ implicite pour les spectres E_i et E_r).

La stimulation des différents photorécepteurs rétiniens par une lumière de com-

position spectrale $E(\lambda)$ peut être représentée par le vecteur $s = \int \rho(\lambda)E(\lambda)d\lambda$, où $\rho(\lambda) = (\rho_1(\lambda), \rho_2(\lambda), \rho_3(\lambda))$ est la fonction vectorielle indiquant la sensibilité des trois types de photorécepteurs (Stockman and T., 2000). En tenant compte de la modulation du flux lumineux par l'ouverture γ de l'iris, il existe donc une dépendance sensorimotrice :

$$\begin{aligned}\varphi(p, \theta, \gamma) &= \gamma \int \rho(\lambda) (\alpha(\theta)E_i(\lambda, p) + \beta(\theta)E_i(\lambda, p)R(\lambda)) d\lambda \\ &= \gamma (\alpha(\theta)s_i(p) + \beta(\theta)s_r(p))\end{aligned}$$

Selon l'approche déjà mise en oeuvre dans le chapitre 6.5, il se pose donc la question de la mise en relation, au travers d'adaptations sensorimotrices, d'une variété de dépendances obtenue par différents environnements. Dans ce qui suit, il est proposé de considérer une variété de dépendances induites par des surfaces ayant les mêmes caractéristiques de spécularité, c'est-à-dire les mêmes fonctions α et β .

Plusieurs adaptations motrices génériquement compensatoires peuvent être mises en évidence. Bien sûr, les rotations ou translation du champ de lumière sont compensables par une adaptation motrice adéquate, c'était l'objet de la section sur l'espace. Mais également, l'adaptation de l'iris $h : (p, \theta, \gamma) \mapsto (p, \theta, k(p)\gamma)$ (déjà évoquée), est compensable par un changement de lumière incidente $E_i(p, \lambda) \mapsto (1/k(p))E_i(p, \lambda)$. Enfin, il en existe d'autres : en particulier, il existe des adaptations génériquement compensatoires qui compensent des changements de réflectance de la forme $R \mapsto kR$, des changements de réflectance de la forme $R \mapsto R + kI$, où $I(\lambda) = 1$ pour tout λ , ou une combinaison des deux. L'idée est que ces différents changements laissent la stimulation $\varphi(p, \theta, \gamma)$ dans le plan engendré par $s_i(p)$ et $s_r(p)$, et qu'il est donc possible de d'obtenir la même stimulation sensorielle, après les changements de réflectance indiqués, en modifiant l'angle d'observation et l'ouverture de l'iris. L'existence de ces adaptations motrices compensatoires apparaît graphiquement sur la figure 9.2.

On peut également noter que seules les dépendances associées aux surfaces achromatiques (i.e. grises, $R = kI$) sont invariantes par certaines de ces adaptations motrices génériquement compensatoires : la réflectance kI est en effet invariante par la famille de changements $R \mapsto k'R + k(1 - k')I$, $k' \in \mathbb{R}$. Ainsi, on voit que les surfaces achromatiques se distinguent des autres propriétés chromatiques par des propriétés particulières d'invariance sensorimotrice.

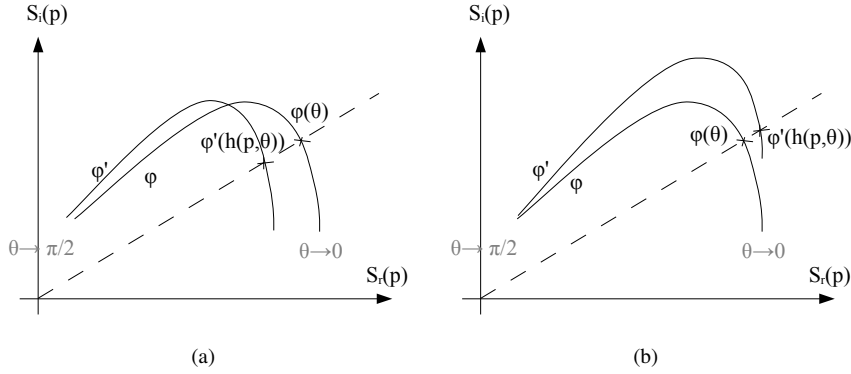


FIG. 9.2 – Adaptation motrice génériquement compensatoire pour les transformations de dépendance associés aux changements de réflectance $R \mapsto kR$ (à gauche) et $R \mapsto R + kI$ (à droite). Dans les deux figures, l'image de la dépendance sensorimotrice φ pour p et γ fixe, c'est-à-dire $\{\varphi(p, \theta, \gamma), \theta \in [0, \pi/2]\}$, est tracée dans le plan engendré par $s_i(p)$ et $s_r(p)$ où ses coordonnées sont simplement (α, β) . L'image de la nouvelle dépendance φ' est dans le premier cas le résultat de la transformation $(\alpha, \beta) \mapsto (\alpha, k\beta)$, et dans le second cas de la transformation $(\alpha, \beta) \mapsto (\alpha + k\beta, \beta)$, et l'on voit que ces transformations peuvent être compensées par des adaptations motrices de la forme $h : (p, \theta, \gamma) \mapsto (p, h(\theta, p), \gamma(\theta, p))$.

Il est intéressant de noter que les adaptations précédentes ne paraissent *pas* permettre de distinguer les adaptations compensatoires des changements $R \mapsto kR$ de celles des changements $R \mapsto kR + (1 - k)I$: si l'on peut distinguer ces deux dernières des autres compensations parce qu'elles caractérisent des séquences de dépendances convergentes vers l'une des extrémités de la variété des dépendances achromatiques, le noir et le blanc sont symétriques vis-à-vis de la structure de ces adaptations.

Outre des adaptations motrices génériquement compensatoires, il existe également dans le modèle dichromatique des adaptations *sensorielles* génériquement compensatoires. Cela est en particulier apparent si l'on se place dans le contexte d'un environnement lumineux naturel. Dans ce cas, une approximation raisonnable pour le champ lumineux est en effet de la forme (Romero et al., 1997; Chiao et al., 2000; Hernandez-Andres et al., 2001) :

$$E_i(\lambda, p) = \sum_{n=1 \dots 3} a_n(p) E_i^n(\lambda)$$

et de simples manipulations d'algèbre linéaires montrent alors que la dépendance sen-

sensorimotrice peut s'écrire :

$$\varphi(p, \theta, \gamma) = \gamma (\alpha(\theta)Id + \beta(\theta)A) [s_i]a(p)$$

où $A = [s_r][s_i]^{-1}$, avec $[s_i]$ et $[s_r]$ des matrices 3×3 correspondant aux stimulations des photorecepteurs pour la lumière incidente et la lumière réfléchie, pour les trois spectres $E_i^1(\lambda)$, $E_i^2(\lambda)$ et $E_i^3(\lambda)$. L'existence d'adaptations sensorielles génériquement compensatoires se voit alors par l'égalité :

$$M\varphi(p, \theta, \gamma) = \gamma (\alpha(\theta)Id + \beta(\theta)MAM^{-1}) [s_i][s_i]^{-1}M[s_i]a(p)$$

autrement dit : l'adaptation sensorielle $s \mapsto Ms$ compense la modification d'environnement caractérisé par les deux changements conjoints $a(p) \mapsto [s_i]^{-1}M[s_i]a(p)$ (changement d'illuminant), et $A \mapsto MAM^{-1}$ (changement de réflectance).

Il s'en suit que les réflectances et les champ lumineux pour lesquels l'adaptation sensorielle $s \mapsto Ms$ laisse la dépendance sensorimotrice invariante sont solutions en M des équations $M[s_i]a(p) = [s_i]a(p)$ (le vecteur $[s_i]a(p)$ est invariant par M) et $MAM^{-1} = A$. Or il est facile de caractériser, par étude de cas, les solutions de cette dernière équation (on rappelle que A est une matrice 3×3), et de voir que le nombre d'invariances de la dépendance sensorimotrice caractérisée par $a(p)$ et A est déterminé par la multiplicité des racines du polynôme caractéristique de A : de nombreux invariants si les trois racines sont égales, moins si seulement deux sont égales, et encore moins si elles sont toutes distinctes. Selon le même argumentaire que celui avancé à propos de la singularité du silence, de l'obscurité, ou des surfaces achromatiques, il y a donc une distinction naturelle à établir entre les différentes propriétés chromatiques des surfaces, selon l'existence d'une plus ou moins grande quantité d'invariances.

Pour conclure cette esquisse d'une approche sensorimotrice de la catégorisation des couleurs, il faut noter que ce qui précède se contente de pointer l'existence d'adaptations motrices et sensorielles, et qu'une question d'intérêt serait bien sûr de caractériser de manière exhaustive, à partir du modèle de dépendance sensorimotrice, l'ensemble des adaptations génériquement compensatoires. Il me semble qu'il ne doit pas en exister d'autre, mais je n'ai pas pu le montrer.

9.2 Article : Color naming...

L'article qui suit ne se place pas explicitement dans l'approche sensorimotrice juste évoquée, mais se trouve naturellement en relation avec elle puisqu'il s'agit de l'étude numérique des valeurs propres des matrices A obtenues pour un ensemble de surfaces artificielles et naturelles¹. Il y est montré que certaines caractéristiques de ces valeurs propres sont en relation étroite avec le coeur des données collectées par la psychophysique sur le manière dont les sujets catégorisent les couleurs.

Il faut noter que ces résultats sont en fait indépendant de l'argumentation précédente, et que l'on pourrait sans doute leur imaginer d'autres argumentation pour les expliquer. En particulier, dans une approche statistique où la problématique serait d'identifier une surface à partir de la stimulation sensorielle produite par la lumière qu'elle réfléchit, étant donnée une loi de probabilité sur son illuminant, il apparaît que la matrice A est un paramètre naturel pour caractériser la loi de probabilité de la stimulation. J'ai tenté d'explorer une telle approche, dont l'idée est d'étudier les propriétés de courbure, pour la métrique de Fischer, de la variété des lois de probabilité de la stimulation obtenues pour un ensemble de réflectance. Mais je n'ai pas réussi à la mener à terme, principalement en raison de difficultés techniques.

Une autre alternative est que nous ayons besoin d'une manière ou d'une autre d'apprendre la relation A pour identifier la propriété de la surface. La vitesse d'apprentissage d'une telle relation par un réseau de neurones dépendant des valeurs propres de cette relation (Cun et al., 1991), les surfaces pour lesquelles la relation A se réduit principalement à une valeur propre seraient apprises plus vite.

Il existe certainement encore d'autres alternatives. Le point ici est simplement de dire que ce qui peut sembler être une argumentation vague dans l'article sur l'importance des valeurs propres de cette matrice peut trouver diverses explications. Il est également de dire que, indépendamment de toute théorie philosophique sur ce qu'est l'expérience sensorielle, d'un pur point de vue physiologique les différentes surfaces utilisées dans les études sur la catégorisation des couleurs ne sont pas équivalentes les unes aux autres vis-à-vis des fonctionnements dans lesquels nous pouvons nous engager, et que ce fait semble fortement corrélé avec les résultats obtenus sur cette catégorisation.

¹L'article se focalise sur le nombre de valeurs propres nulles plutôt que sur le nombre de valeurs propres égales, mais voir les "supplementary informations" qui le suivent : en fait les deux critères conduisent à des résultats très similaires.

Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties

DAVID L. PHILIPONA AND J. KEVIN O'REGAN

Laboratoire de Psychologie de la Perception, CNRS, Université Paris 5 René Descartes, Paris, France

(RECEIVED August 1, 2005; ACCEPTED December 29, 2005)

Abstract

Psychophysical studies suggest that different colors have different perceptual status: red and blue for example are thought of as elementary sensations whereas yellowish green is not. The dominant account for such perceptual asymmetries attributes them to specificities of the neuronal representation of colors. Alternative accounts involve cultural or linguistic arguments. What these accounts have in common is the idea that there are no asymmetries in the physics of light and surfaces that could underlie the perceptual structure of colors, and this is why neuronal or cultural processes must be invoked as the essential underlying mechanisms that structure color perception. Here, we suggest a biological approach for surface reflection properties that takes into account only the information about light that is accessible to an organism given the photopigments it possesses, and we show that now asymmetries appear in the behavior of surfaces with respect to light. These asymmetries provide a classification of surface properties that turns out to be identical to the one observed in linguistic color categorization across numerous cultures, as pinned down by cross cultural studies. Further, we show that data from psychophysical studies about unique hues and hue cancellation are consistent with the viewpoint that stimuli reported by observers as special are those associated with this singularity-based categorization of surfaces under a standard illuminant. The approach predicts that unique blue and unique yellow should be aligned in chromatic space while unique red and unique green should not, a fact usually conjectured to result from nonlinearities in chromatic pathways.

Keywords: Color perception, Reflection properties, World color survey, Unique hues, Hue cancellation

Introduction

Particular hues of red, green, blue, and yellow are considered as “pure colors”, often used in color-classification schemes (Hering, 1920/1964; Boynton & Olson, 1987; Chichilnisky & Wandell, 1999). There are strong indications that the choice of these colors is not a mere cultural convention since these same colors are consistently singled out and given names even in widely separated human societies (Berlin & Kay, 1969; Regier et al., 2005). Furthermore, psychophysical tests provide measures of the fact that these colors appear unmixed, or “unique”, in the sense that they seem to observers not to be tinted with any other color (Jameson & Hurvich, 1955; Valberg, 2001).

These facts, which are considered to be at the heart of color vision, have up till now not been adequately explained.

First, there is no known *physical* reason why certain colors should be perceived as unmixed or special. In particular, lights that elicit sensations judged to be unmixed by observers can actually be

obtained by superposing other lights, and paints judged to be unmixed can be obtained by mixing other paints.

Second, no accurate *biological* basis for unique hues has so far been found. The trichromatic theory of human vision, based on three types of photoreceptors, would seem unable to explain the existence of exactly four unique hues. The opponent-process theory argues that there are four perceptually unique hues because of the existence of so-called red/green and blue/yellow neural pathways, but specialists in color vision agree that when the theory is confronted with available physiological data, it does not adequately predict which hues are observed to be perceptually unique (Knoblauch et al., 1985; Webster et al., 2000; Valberg, 2001; Kuehni, 2004).

Here we show theoretically and numerically that, contrary to what is usually thought, the existence of four special surface colors can be expected from trichromatic theory alone, when constraints satisfied by natural illuminants and surfaces are taken into account. The colors predicted to be special from this approach are precisely those which are found in empirical surveys to be most often given a name by different cultures throughout the world. We shall additionally see that the approach provides correct quantitative predictions for foundational facts of color science, namely psychophysical data about unique hues and hue cancellation.

Address correspondence and reprint requests to: J. Kevin O'Regan, Laboratoire de Psychologie de la Perception, Université Paris 5 René Descartes, 45 rue des St. Pères, 75006 Paris, France. E-mail: oregan@ext.jussieu.fr

Terminology: accessible information

In this paper, we will call “accessible information” about a light the restricted information about the spectral composition of that light which is accessible through an organism’s photopigment set: For humans, we can take for instance the three numbers corresponding to the photon absorption counts for each of the photopigment types present in human photoreceptors. We use the term “accessible information” rather than “spectral coding” or “photoreceptor excitations” because the two latter terms leave an ambiguity as to whether they refer to the number of photons absorbed by photopigments, or the neural coding of these numbers. We want to be clear that the constraint we analyze herein really is about photon catches and not neural code for color information: in particular, properties of linearity that we consider refer to catches, not to photoreceptor outputs.

A biological approach for reflection properties

The approach we propose involves constructing a biological restriction of the physicist’s notion of reflectance. From the physicist’s viewpoint, reflectance is the relationship that holds between the spectrum of the light illuminating a surface and the spectrum of the light reflected by that surface. To take into account the loss of information induced by photopigments, it is therefore natural for the biologist to study the relationship between the accessible information (instead of the spectrum) about the illuminant and the accessible information (instead of the spectrum) about the reflected light. We will show that, with respect to standard sets of natural illuminants, physical surfaces’ reflectance induce a linear constraint between the accessible information about the illuminant and the accessible information about the light reflected by that surface. This suggests a biological conception of surface reflection properties as *linear operators* in the three-dimensional space of the accessible information.

The work of the paper will consist in studying the characteristics of these linear operators for a large number of natural and artificial surfaces. We shall show that for certain surfaces the operator is *singular*: instead of variations in the incoming light causing variations of the reflected light along all *three* directions of the accessible information space as is usually the case, in the singular cases, variations in incoming light only produce strong variations in the reflected light along one or along two directions. These singular cases where a surface affects incoming light in a simpler way than usual will be seen to predict, with surprising accuracy, data on color naming, unique hues, and hue cancellation.

Databases and numerical methods

Database for photopigments

We used the 10-deg Stiles and Burch Color-Matching Functions (CMFs), arguably the most reliable experimental data about the *in situ* sensitivity of the set of human photopigments (Stiles & Burch, 1959). Since our approach does not rely on a linear basis for the accessible information, we do not actually need to go back to the absorption of individual photopigments: we could simply perform our numerical analysis directly from CMFs. We checked that using Stockman and Sharpe (2000) cone fundamentals does indeed not bring any noticeable modification of the results. In the theoretical derivations below, we refer to photopigment absorptions since these are the original physical constraints.

Database for illuminants

We used a set of illuminants composed from 99 daylight spectra in Granada from Romero et al. (1997), 238 daylight spectra in forest from Chiao et al. (2000), and a Gaussian sample of 200 spectra constructed from basis functions S0, S1, and S2 derived by Judd et al. (1964). The rationale was to consider diverse sources of illuminants so as to avoid possible statistical artifacts resulting from considering a too specific ecological niche.

Database for reflectances

We used two sets of reflectances: one made from artificial surfaces, and one made from natural surfaces. The set of artificial surfaces was constituted by Munsell chips, whose reflectances were taken from the University of Joensuu (Parkkinen et al., 1989—<http://spectral.joensuu.fi/>). We chose the set of 1600 Munsell glossy chip reflectances because it is the one from which chips used for the World Color Survey were taken. The set of natural surfaces comprised 404 reflectances of leaves, petals, and barks from the grounds of Keele University in Staffordshire (Westland, Shaw, & Owens, 2000), 1862 reflectances of fruits and foliages from French Guiana and Uganda, made available by the University of Cambridge, 1413 reflectances of flowers and 246 reflectances of leaves from the Chittka dataset (Chittka & Menzel, 1992).

Numerical methods

All numerical computations were performed using Matlab 7 (Mathworks, Natick, MA). All spectra were resampled for compatibility at 5-nm intervals within the 400–700 nm band of the spectrum. Numerical estimation of integrals was performed by a simple dot product between sampled spectra. Standard algebraic computations such as multivariate regression and decomposition in eigenvalues and eigenvectors were performed using standard Matlab routines without any particular options.

Mathematical treatment

In this section, we first verify numerically that over our datasets surface reflectances can accurately be described by linear operators on accessible information. We then introduce the concept of noninteracting basis for accessible information and show how it can be used to define a singularity index for each surface in the databases. The Results section confronts the measured singularity for each surface to existing psychophysical data on color naming, unique hues, and hue cancellation.

Computing biological reflectance mapping

Consider an illuminated surface. Let $E(\lambda)$ denote the spectral power distribution of the light incident on the surface at each wavelength λ . Let $R_1(\lambda)$, $R_2(\lambda)$, and $R_3(\lambda)$ denote the absorption rate at each wavelength λ by photopigments present in the L, M, and S human photoreceptors. Then the accessible information about the incident illuminant can be described by the triplet $u(E) = (u_1(E), u_2(E), u_3(E))$, where

$$u_i(E) = \int R_i(\lambda) E(\lambda) d\lambda, \quad i = 1 \dots 3. \quad (1)$$

Let $S(\lambda)$ denote the reflectance function of the surface, then the accessible information about the reflected light is the triplet $v^S(E) =$

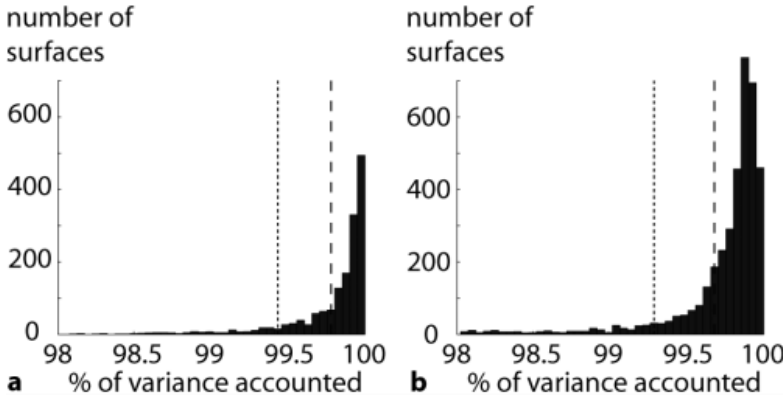


Fig. 1. Variance residuals for the set of Munsell chips (left) and natural surfaces (right). Plots show the number of surfaces as a function of the variance of the accessible information for the reflected light accounted for by a linear transformation of the accessible information for the incident light. Dashed line represents the mean value of the variance accounted for over the dataset, dotted line represents the value of the variance at least accounted for by 90% of surfaces.

($v_1^S(E), v_2^S(E), v_3^S(E)$) (superscript S specifies that this information depends on the surface S considered, in addition to the illuminant E) where

$$v_i^S(E) = \int R_i(\lambda) S(\lambda) E(\lambda) d\lambda, \quad i = 1 \dots 3. \quad (2)$$

Now, instead of a unique illuminant assume a probability distribution over illuminants. Then, for a given surface S , the existence of a linear relationship between three-dimensional variables u and v^S , that is, the existence of a linear operator A^S in the accessible information space such that $v^S = A^S \cdot u$, can be assessed by computing the empirical variance of the residual $\eta = v^S - A^S \cdot u$, where A^S is the 3×3 matrix obtained by a linear regression of v^S onto u computed from a sample of illuminants. For each surface of our dataset, we therefore performed a linear regression from a series of calculations simulating the illumination of the surface successively by each illuminant of our database of daylights.

Fig. 1 summarizes the variance of residuals obtained. Results show that the variance left after taking into account a linear relationship between u and v^S is negligible: in terms of average value over our two sets of surfaces, this residual represents 0.3% of the original variance of v^S for the set of natural surfaces and 0.2% for the set of Munsell chips, and in terms of quantiles, more than 90% of both natural and artificial surfaces have a residual smaller than 0.7% of the original variance of v^S . This confirms that for the purpose of color vision, we can restrict our view of physical reflection properties to a biological conception of linear operators in the three-dimensional space of the accessible information.

This result was expected from the well-known low dimensionality of the set of natural illuminant spectra. The 400–700 nm band of daylight spectra are well approximated by weighted sums of only a few functions, presumably as few as three (Judd et al., 1964; Maloney, 1999; Chiao et al., 2000). In these conditions, let $E_1(\lambda)$, $E_2(\lambda)$, $E_3(\lambda)$ be such basis functions for illuminant spectra and ε_1 , ε_2 , ε_3 be the coordinates of $E(\lambda)$ in this basis so that

$$E(\lambda) = \sum_{k=1}^3 \varepsilon_k E_k(\lambda). \quad (3)$$

Build 3×3 matrices $U = [u(E_1), u(E_2), u(E_3)]$ and $V^S = [v^S(E_1), v^S(E_2), v^S(E_3)]$, then we have $v^S(E) = V^S(\varepsilon_1, \varepsilon_2, \varepsilon_3)^T$ and $u(E) = U(\varepsilon_1, \varepsilon_2, \varepsilon_3)^T$, therefore if U is invertible (basis functions yield linearly independent photon catches triplets), we find that $v^S(E) = V^S U^{-1} u(E)$. Thus, the three dimensionality

of the illuminant space implies a linear relationship between u and v^S .

Note that the linear operator $A^S = V^S U^{-1}$ does not depend on a probability distribution of illuminants within the linear model of eqn. (3), but recall also that the illuminant space spanned by such a model is itself determined by a probability distribution over the set of all possible illuminants. Note finally that even though it provides an intuitive rationale, the three dimensionality of the illuminant space (questioned for instance by the work of Romero et al. (1997)), is not mandatory to find a linear relationship between u and v^S since some aspects of the spectrum to be modeled by basis functions might have little influence on both u and v^S .

Noninteracting bases for accessible information

Since the biological analog A^S of a physical reflectance $S(\lambda)$ is a simple linear operator in the accessible information space, we can express it in terms of eigenvalues and eigenvectors, the possibility of such a decomposition in the complex field relying on a rather weak assumption (see Appendix on real and complex eigenvalues). If the decomposition exceptionally turns out to occur in the real field, then eigenvalues and eigenvectors have straightforward interpretation. Indeed, the three eigenvectors of A^S will provide a basis for the accessible information space such that elements of this basis do not mutually interact with each other when they are reflected by the surface $S(\lambda)$, and are simply individually scaled by the associated eigenvalues.

It is very simple in this basis to determine, for the surface S , the accessible information about the reflected light from the accessible information about the illuminant: let $\{r_1^S, r_2^S, r_3^S\}$ and $\{C_1^S, C_2^S, C_3^S\}$ be the three eigenvalues and three eigenvectors of the operator A^S , and assume that $u(E)$ has coordinates $(\alpha_1^S, \alpha_2^S, \alpha_3^S)$ in the basis* $\{C_1^S, C_2^S, C_3^S\}$, then

$$\begin{aligned} u(E) &= \sum_{j=1}^3 \alpha_j^S C_j^S \Rightarrow v^S(E) = A^S u(E) \\ &= \sum_{j=1}^3 \alpha_j^S A^S C_j^S \\ &= \sum_{j=1}^3 r_j^S \alpha_j^S C_j^S \end{aligned} \quad (4)$$

*Note that although $u(E)$ is not dependent on the surface considered, the set of eigenvectors of A^S is, and therefore the coordinates of $u(E)$ in that basis do depend on the surface.

that is, $v^S(E)$ has coordinates $(r_1^S \cdot \alpha_1^S, r_2^S \cdot \alpha_2^S, r_3^S \cdot \alpha_3^S)$ in this same basis. This also means that it is easy in this basis to discount the illuminant from the accessible information about the reflected light: it suffices to divide independently each coordinate of the accessible information about the reflected light by the same coordinate of the accessible information about the illuminant. Note that this is only possible in this basis, and therefore *a fortiori* only possible if A^S has real eigenvalues.

It is a remarkable fact that noninteracting bases can be found for most physical surfaces: in fact the majority of linear operators associated with reflectance functions of our dataset turn out to be decomposable in the real field. 88% of surfaces over our artificial and natural datasets have real eigenvalues, and for the remaining surfaces the bound $\rho \sin(|\theta|/2)$ (see Appendix on real and complex eigenvalues) never exceeds 0.023 for the set of Munsell chips and 0.012 for natural surfaces. Therefore, it is possible for each surface to find a noninteracting, or approximately noninteracting, basis $\{C_1^S, C_2^S, C_3^S\}$ and real reflection coefficients $\{r_1^S, r_2^S, r_3^S\}$. The point is that these bases are different from surface to surface. Fig. 2 shows examples of light changes associated with such bases, for two different surfaces.

The three reflection coefficients defined in this way for each surface then almost provide a biological analog of the physicist's reflectance function. Indeed, just as the physicist describes the reflectance function in terms of how strongly a surface reflects each of an infinity of noninteracting monochromatic light components, we here can describe a biological analog in terms of the three numbers which define how strongly the surface reflects each of the three noninteracting components of the accessible information. It must be noted that coefficients $\{r_1^S, r_2^S, r_3^S\}$ do not depend for instance on the fact that we are using color-matching functions rather than cone absorption curves to *represent* the accessible information. Indeed, CMFs are essentially a linear combination of cone absorption curves (Wandell, 1997), and therefore using them rather than cone absorption curves merely leads to a linear change of coordinates for the accessible information. Since we did not use any *a priori* particular coordinate system to define noninteracting basis and reflection coefficients, this has no influence on reflection coefficients.

Whereas reflection coefficients are an analog of the physicist's reflectance function, the analogy is not complete because in contrast with the physical basis of monochromatic spectra and as already stressed, the biological basis may vary from surface to surface (hence the superscript S in C_i^S). Yet the existence of a unique basis actually is an implicit assumption of several standard approaches to color constancy. The accessible information for the illuminant being the accessible information for the light reflected by a white surface, the existence of a unique basis for a set of

surfaces is, in addition to the gray world hypothesis, the condition for a von Kries algorithm in this basis to achieve color constancy of a scene constituted by such surfaces. From this viewpoint, our work relates to previous work on spectral sharpening by Finlayson et al. (1994). As for cone ratios between surfaces in a scene (Nascimento et al., 2002), their rigorous invariance can be seen to require a unique, and additionally very specific, basis for these surfaces: the basis defined by cone catches. In this paper, however, we will not elaborate on the existence of a unique basis for a set of surfaces and stick to the investigation of the biological constraints induced by each surface *individually*.

Singular reflection properties

Contrary to the situation in physics where there are an infinity of reflection coefficients, the biological case, with only three coefficients, presents asymmetries determined by a few specific configurations in the magnitudes of these coefficients. A first such singular case is the case when all three reflection coefficients are about equal, that is, when the three noninteracting components of incoming light are equally transmitted. This case clearly corresponds to achromatic surfaces, since the spectral composition (or rather, what is biologically accessible about this composition) remains unchanged by the surface. Examples of such cases will be white surfaces and black surfaces, with all three reflection coefficients equal to 1 or to 0, respectively.

Another type of singularity will consist in cases when variations in incoming light only produce strong variations in the reflected light along one or along two directions. Like black and white, we also expect that these cases will have a special perceptual status. To quantify the extent to which a surface approaches one of these two special cases, we considered the two ratios $\beta_1 = r_1/r_2$ and $\beta_2 = r_2/r_3$, where (r_1, r_2, r_3) are the reflection coefficients sorted in decreasing order. Maxima of β_1 correspond to surfaces that reflect one of the elements of the noninteracting basis much more significantly than the two others; maxima of β_2 correspond to surfaces that reflect two of elements much more significantly than the third. Because we would like for simplicity to have a unique index summarizing these two cases, we could consider just taking the maximum of β_1 and β_2 . But it turns out that for our dataset of surfaces, β_1 peaks at about 7 while β_2 peaks at about 21, hence such an index would essentially pinpoint how close we are from the second case. To deal with this issue, we first normalized each of β_1 and β_2 to their maximum value so that value 1 means for both "maximum singularity", and only then gathered the two ratios into a single value by taking their maximum. To sum up, we defined our singularity index as $\beta = \max(\beta_1/\beta_1^{\max}, \beta_2/\beta_2^{\max})$, where β_1^{\max} and β_2^{\max} are the maximum values of each β value over the dataset of surfaces.

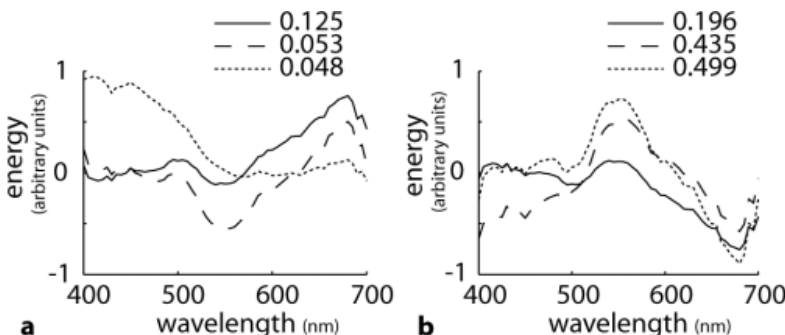


Fig. 2. Illuminant changes associated with a noninteracting basis of the accessible information: (a) Munsell chip 10 R 2/6, (b) Munsell chip 5 BG 7/8. The accessible information for these components of the illuminant is merely scaled by the surface, and by the amount specified in the legend.

Results

The Mathematical treatment section has shown that there is for each surface a linear operator which describes the way the surface transforms accessible information about illuminating light to accessible information about reflected light. We have also shown how to define a *singularity index* which quantifies the degree to which this linear operator corresponds to the special cases where the variation of reflected light, under changes in illuminant, is confined essentially to only one or two component directions, rather than being distributed over three component directions, as is normally the case. We now confront the singularity index, computed over large databases of surfaces with empirical data on color naming, unique hues and hue cancellation. We stress that the singularity index was calculated from existing databases of illuminants, photoreceptor fundamentals, and illuminants, without parameter adjustment of any kind.

Singular reflection properties and the World Color Survey

Fig. 3b plots the value of the singularity index for a standard set of surface patches, namely the colored chips used to study color

categorization in cross-cultural studies (the World Color Survey, Berlin & Kay, 1969, made available by the International Computer Science Institute, Berkeley). The chips most often given a name by widely separated human cultures (Fig. 3a), and which we call “red”, “yellow”, “green” and “blue” in English, can be seen to be within one chip of those having maximally singular reflecting properties. According to the index we have defined, these cases are reached for chips labeled G2, C9, F16, and H31, with reflection coefficients $\{0.56, 0.08, 0.08\}$, $\{0.89, 0.66, 0.03\}$, $\{0.29, 0.07, 0.04\}$, and $\{0.27, 0.07, 0.07\}$.

Two striking facts should be noted from Fig. 3. First, the figure shows that there is a biological basis for the fact that precisely four categories of colored surfaces should be perceived as distinctive: these are singular in the sense that instead of reflecting three components, they reflect essentially only one or two components of the light described by biologically accessible information. The fact that a finite number of categories for reflection properties emerge from an infinity of possible reflectances, and further that four categories emerge from taking into account a trichromatic constraint on the visual system, already is a notable result from this biological reading of reflection properties. A second notable result from Fig. 3 is that these four categories of surfaces precisely

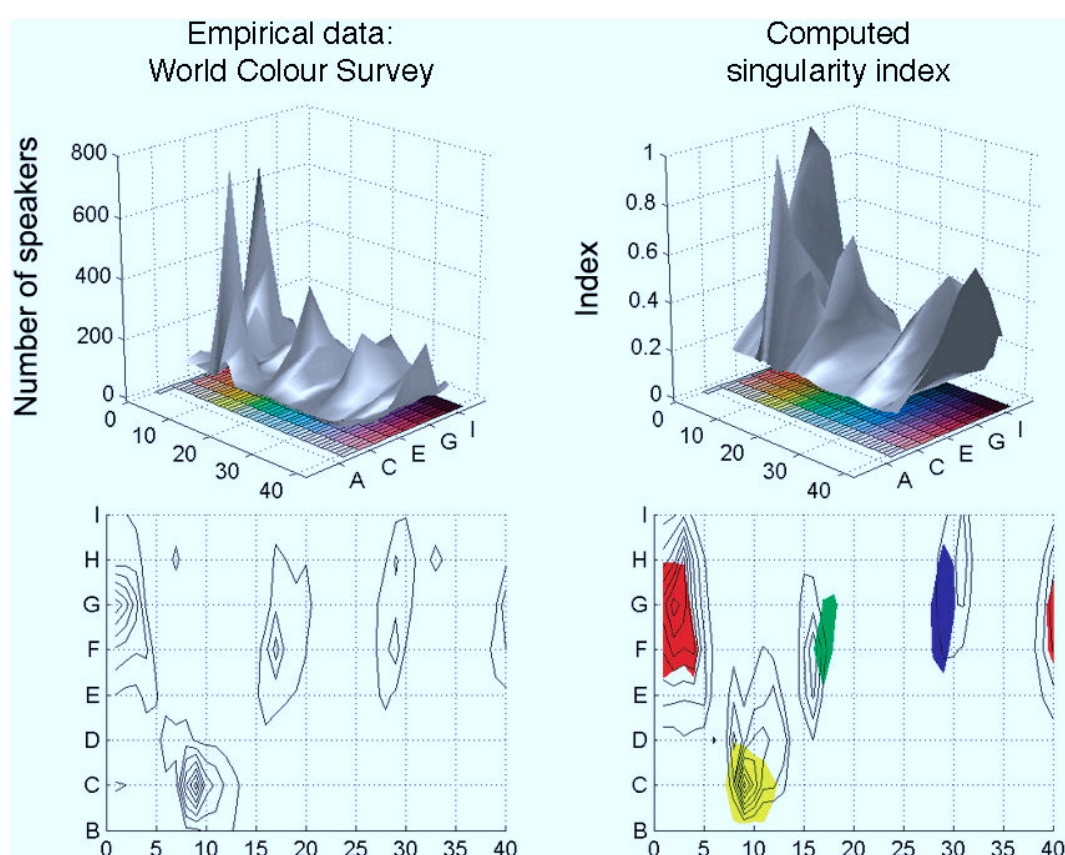


Fig. 3. Color terms across languages, and surface reflecting properties. Left panels show the number of speakers in Berlin and Kay’s World Color Survey of 110 languages having a term that designates each of a set of colored surfaces. There are four main maxima, corresponding to four universal “focal colors”: red, yellow, green, and blue. Right panels show our theory-based calculation for the maximum of the two ratios defined by the first vs. second, and second vs. third reflection coefficients (see text) for these surfaces. For comparison with the empirical data, colored areas in the bottom right panel correspond to those chips in the World Color Survey that were named by more than 20% of the maximum number of speakers. Letter and number notations for chips are those used in the World Color Survey.

correspond to colors most often given a name across widely separated cultures, namely red, yellow, green, and blue.

Both these facts, not previously noticed, have been obtained without appealing to neural mechanisms that underlie opponent channels in the visual system. They are merely a consequence of the asymmetries in surface reflecting properties induced by the limited way biological photoreceptors sample physical spectra. It could thus be argued that the reason the colors “red”, “yellow”, “green”, and “blue” are so often singled out among all other colors as being worth giving a name, is that surfaces of these colors have the particularity that they alter incoming light in a simpler way than other surfaces: red, green, and blue surfaces reflect essentially one component. Yellow surfaces are also simple as compared to other surfaces in that they reflect two components.

Singular reflection properties and unique hues

The coincidence between singularities in reflecting properties and cross-cultural linguistic color categories suggests that beyond accounting for anthropological results, singularities in reflecting properties might also be correlated with more precise data collected by psychophysicists about color perception. To investigate this, we need to link our results about colored surfaces under natural illuminants to psychophysical results where observers classically face “aperture colors”. Such stimuli are generated, roughly speaking, by sending lights of controlled spectral composition directly into the eye rather than by natural viewing of a colored surface. To compare our approach with the results of these experiments, we made the minimal conjecture that the nervous system interpreted such stimuli as being the result of reflection from a colored surface of light deriving from the most common illuminant type, known as D65. This provides a simple way to transpose the index concerning surface properties to an index concerning photon capture by human photopigments.

To simplify the representation of this transposed index, we reasoned that although it is known that natural reflectances are not accurately described by using sums of only three basis functions, such models nonetheless represent a first approximation that suggests that the index should be roughly invariant by scalar multi-

plication of photon catches. Indeed, if three-dimensional models of reflectance were exact, then the transposed indexes for proportional photon catches would be the indexes of proportional reflectance functions, and hence would be equal since we consider *ratio* of eigenvalues. As a result, the transposed index is expected to vary essentially on a projective structure, and we can plot the index using (x, y) CIE 1931 chromatic coordinates (Wandell, 1997).

Fig. 4a shows the transposed index based on the set of Munsell chips. For natural surfaces on the other hand, plotted in Fig. 4c, we found (as noted elsewhere, see for instance Westland et al., 2000) that the gamut spanned in chromaticity space by natural surfaces was smaller than that spanned by Munsell chips, and in particular too small in the green area to get a readable plot. We therefore artificially saturated the reflectances by generating a reflectance function $\tilde{S}(\lambda) = S(\lambda) - \min S$ for each natural reflectance of our database. While this manipulation must be kept in mind when reading Fig. 4c, it seems unlikely that the correspondence found then with Fig. 4a is a result of this mere saturation.

The crests of the index in these figures pinpoint stimuli associated with the most singular colored surfaces. It is a remarkable fact that these loci can be fitted by four half-planes in the full space of the accessible information that share as an edge the central axis defined by achromatic surfaces under the hypothesized illuminant, in this case D65 (Fig. 4b). These half-planes determine specific wavelengths at their intersections with the monochromatic locus, and therefore identify the wavelengths producing stimuli that are most exemplary of singular surface reflecting properties. This provides a prediction for the monochromatic lights that we might expect subjects will report as being “unique hues” in psychophysical experiments. As seen in the figure, the predictions are very close to observed empirical data on unique hues. Our approach is thus able, without any parameter adjustments, to predict psychophysically measured unique hues purely from known biological and physical data (photopigment absorption curves, and illuminant spectra and reflectance functions, respectively).

While the hypothesized illuminant, D65 in Fig. 4, could be suspected to play an essential role in the predictions of unique hue wavelength in this approach, this conjecture turns out to be wrong, at least for the case of red, green, and blue: this is because these

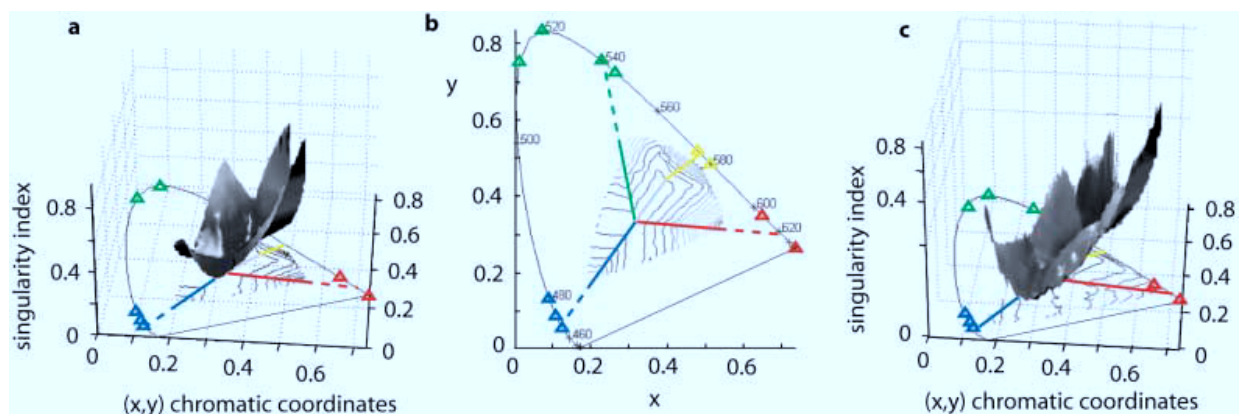


Fig. 4. Surface reflecting properties and unique hues. (a) Singularity index computed from the set of Munsell chips under daylight D65, plotted as a function CIE chromatic coordinates. Crests of the index are projected onto the ground plane as solid colored lines, marking the loci of maximum singularity. (b) Crests of the index extended out to the monochromatic locus. Colored triangles: psychophysical measures of unique hues from several sources given in Table 1. (c) Singularity index computed from a set of natural surfaces under daylight D65, plotted as a function CIE 1931 chromatic coordinates.

surfaces reflect only one component of the accessible information about the incident light and therefore the accessible information about the reflected light varies little in chromaticity as the hypothesized illuminant changes. It turns out numerically that the wavelength of unique yellow is also not strongly dependent either on the temperature of the illuminant: as the illuminant hypothesized to interpret tristimuli varies from temperature 3000°K to 20,000 °K, the wavelength of the unique yellow determined by our approach varies from 580 nm to 570 nm.

Another issue concerns the variability of observed unique hues. While there is evidence that visual experience plays an important role in the perception of unique hues as compared to phylogenetic neuronal specifics (Brainard et al., 2000; Lotto & Purves, 2002; Neitz et al., 2002; Sugita, 2004), it has until now been unclear what features of visual experience could influence, or arguably determine, unique hues. A rough indication about the expected variability of a unique hue within our approach, under the assumption that the singularity is relevant if it is above some threshold, can be given by the span over which the singularity index is “large”, for instance greater than half of its value for the unique hue. Table 1 shows a comparison of these predictions with empirically observed unique hues and demonstrates that the specific pattern of variability found between observers (Kuehni, 2004) for the different unique hues is in agreement with what is expected from our approach. This is further support that singularities in the biologically accessible surface reflectance properties may indeed constitute the relevant features for understanding unique hues.

A last point concerns the well-established fact that unique yellow and blue are aligned in chromaticity space, while unique red and green are not. This is usually considered to result from a quirk of the neural system’s structure (Chichilnisky & Wandell, 1999; Valberg, 2001; Knoblauch & Shevell, 2001), while our work suggests that it could rather reflect a physical fact about the information processed by the visual system: as seen from Fig. 4, the crest in the singularity index for red and green are almost colinear, whereas this is clearly not the case for blue and yellow.

Singular reflection properties and hue cancellation

A further consequence of the findings presented in this paper relates to the well-known phenomenon of hue cancellation. Hue cancellation quantifies the fact that the addition to a light that appears bluish of a certain amount of light that appears yellowish produces a light that appears neither bluish nor yellowish, and the same for lights that appear reddish or greenish. This has been considered as evidence that color perception was mediated by two sensation continua, redness–greenness and yellowness–blueness (Jameson & Hurvich, 1955; Valberg, 2001). Yet it is a fact, noted for instance by Mollon and Jordan, 1997, that hue-cancellation data are a direct consequence of the loci of unique hues: in Fig. 5a, the dash–dot cancellation curve were derived from computing the intensity of a monochromatic yellow light to add to a bluish light so that the corresponding stimulus is on the locus defining a unique hue different from yellow or blue (or the achromatic locus), and similarly for the dashed curve with red and green lights.

Since we have an account of the loci of unique hues from singular surface reflection properties, we can make predictions about hue cancellation without appealing to equilibration of “sensation channels” (Fig. 5b). Thus, we argue that hue-cancellation data cannot be considered as conclusive evidence for a Müllerian theory of two sensation continua. Obviously, this does not deny the importance of color-opponent neuronal pathways for color vision: in particular, they optimize information transfer from the retina (Buchsbäum & Gottschalk, 1984; Ruderman et al., 1998), and explain aftereffects (McCollough, 1965). But it casts doubt on the claim that this neuronal specificity on its own straightforwardly determines a fact about our conscious perceptual experience: the existence of a few special hues.

Conclusion

We have shown how by considering the information accessible *via* human photoreceptors about incoming light, it is possible to

Table 1. Comparison between psychophysical measurements of unique hues (Kuehni, 2004) and predictions based on the structure of the singularities of surface reflecting properties. EOS means End Of Spectrum.

Data set	Subjects	Unique yellow		Unique green	
		Mean (nm)	Range (nm)	Mean (nm)	Range (nm)
Scheffrin	50	577	568–589	509	488–536
Jordan & Mollon	97	—	—	512	487–557
Volbrecht	100	—	—	522	498–555
Webster (a)	51	576	572–580	544	491–565
Webster (b)	175	580	575–583	540	497–566
Webster (c)	105	576	571–581	539	493–567
Prediction	—	575	570–580	540	510–560
Data set	Subjects	Unique blue		Unique red	
		Mean (nm)	Range (nm)	Mean (nm)	Range (nm)
Scheffrin	50	480	465–495	—	—
Jordan & Mollon	97	—	—	—	—
Volbrecht	100	—	—	—	—
Webster (a)	51	477	467–485	EOS	—
Webster (b)	175	479	474–485	605	596–700
Webster (c)	105	472	431–486	EOS	—
Prediction	—	465	450–480	625	590–EOS+

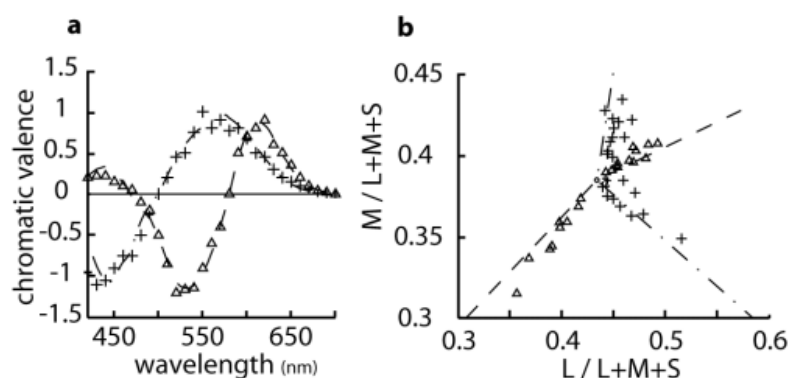


Fig. 5. Prediction of hue cancellation and equilibrium loci. (a) Classic empirical data from Jameson and Hurvich, 1955: Crosses show the relative intensity of yellow (negative ordinate) or blue (positive ordinate) lights added to monochromatic lights of different wavelength so that the observer judged the stimulus as neither yellow or blue. Triangles present similar data for red and green lights. Curves: predictions based on the observer's monochromatic unique hues. (b) Crosses show yellow–blue equilibrium, triangles show red–green equilibrium for one observer, from Chichinisky and Wandell, 1999. Dashed lines: loci for unique hues predicted from singular reflection properties.

characterize reflecting surfaces by a *singularity index*. This index quantifies the degree to which the surface reflects light in a simpler way than usual. Whereas usually a surface will cause variations in accessible information about reflected light along all *three* directions of the accessible information space, singular surfaces only produce strong variations along one or two directions.

We showed how the singularity index, as calculated over a large database of surfaces and illuminants, correctly predicts which surfaces will be given names across widely separated cultures. We then showed how the singularity index can be applied to lights instead of surfaces in order to predict empirically observed unique hues. Data on hue cancellation is therefore also well explained by this approach.

In addition to the surprisingly accurate predictions of classic psychophysical data, made without any parameter adjustments, the approach proposed here has interesting philosophical implications. Colors are a genuine quandary to science because they forcibly instantiate the difficulty of explaining the quality of perceptual awareness. While the dominant view today is that perceptual qualities are determined by specificities of cortical representations (Valberg, 2001; Crick & Koch, 2003; Zeki, 2003), a different stance about the role of nervous activity can be taken. Rather than conjecturing that nervous activity determines the quality of sensations, a conjecture that opens up an “explanatory gap” (Levine, 1983; Chalmers, 1996), we propose instead to exploit the less controversial fact that this neural activity enables the sensorimotor involvement of the organism and its cognitive abilities. What determines the perceived quality of sensations, we suggest then, are intrinsic features of the overall constraints imposed on the interaction of the organism with its environment (O'Regan & Noë, 2001). Colors have for long been used as an exemplary counter-argument to such functional accounts of sensations, yet the common rationale and surprisingly accurate quantitative account provided here for color naming, unique hues and hue cancellation appears to lend credence to this approach.

Acknowledgments

David L. Philipona was supported by the European Network of Excellence Enactive, IST-2004-002114, and the European Project CoSy, FP6-004250-IP. The authors thank K. Knoblauch, S. Dehaene, and J. Clark for inspiring remarks, as well as the two reviewers who greatly helped improving the paper. The authors are particularly indebted to O. Coenen who provided financial and research support for D.L.P. while he was working on this project at Sony CSL.

References

- BERLIN, B. & KAY, P. (1969). *Basic Color Terms: Their Universality and Evolution*. Berkeley, California: University of California Press.
- BOYNTON, R.M. & OLSON, C.X. (1987). Locating basic colors in the OSA space. *Color Research and Application* **12**, 94–105.
- BRAINARD, D.H., ROORDA, A., YAMAUCHI, Y., CALDERONE, J.B., METHA, A., NEITZ, M., NEITZ, J., WILLIAMS, D.R. & JACOBS, G.H. (2000). Functional consequences of the relative numbers of L and M cones. *Journal of the Optical Society of America A* **17**, 607–614.
- BUCHSBAUM, G. & GOTTSCHALK, A. (1984). Chromaticity coordinates of frequency-limited functions. *Journal of the Optical Society of America* **1**, 885–887.
- CHALMERS, D. (1996). *The Conscious Mind*. Oxford: Oxford University Press.
- CHIAO, C.-C., CRONIN, T.W. & OSORIO, D. (2000). Color signals in natural scenes: Characteristics of reflectance spectra and effects of natural illuminants. *Journal of the Optical Society of America A* **17**, 218–224.
- CHICHILNISKY, E. & WANDELL, B.A. (1999). Trichromatic opponent-color classification. *Vision Research* **39**, 3444–3458.
- CHITTKA, L. & MENZEL, R. (1992). The evolutionary adaptation of flower colors and the insect pollinators' color vision systems. *Journal of Comparative Physiology A* **171**, 171–181.
- CRICK, F. & KOCH, C. (2003). A framework for consciousness. *Nature Neuroscience* **6**, 119–126.
- FINLAYSON, G., DREW, M. & FUNT, B.V. (1994). Spectral sharpening: Sensor transformations for improved color constancy. *Journal of the Optical Society of America A* **11**, 1553–1563.
- HERING, E. (1920/1964). *Outlines of a Theory of the Light Sense*. Cambridge, Massachusetts: Harvard University Press.
- JAMESON, D. & HURVICH, L.M. (1955). Some quantitative aspects of an opponent colors theory. I. Chromatic responses and spectral saturation. *Journal of the Optical Society of America* **45**, 546–552.
- JUDD, D.B., MACADAM, D.L. & WYSZECKI, G. (1964). Spectral distribution of typical daylight as a function of correlated color temperature. *Journal of the Optical Society of America* **54**, 1031–1040.
- KNOBLAUCH, K., SIROVICH, L. & WOOTEN, B.R. (1985). Linearity of hue cancellation in sex-linked dichromacy. *Journal of the Optical Society of America A* **2**, 136–146.
- KNOBLAUCH, K. & SHEVELL, S.K. (2001). Relating cone signals to color appearance: Failure of monotonicity in yellow/blue. *Visual Neuroscience* **18**, 901–906.
- KUEHNI, R.G. (2004). Variability in unique hue selection: A surprising phenomenon. *Color Research and Application* **29**, 158–162.
- LEVINE, J. (1983). Materialism and Qualia: The Explanatory Gap. *Pacific Philosophical Quarterly* **64**, 354–361.
- LOTTO, B.R. & PURVES, D. (2002). A rationale for the structure of color space. *Trends in Neuroscience* **25**, 84–88.
- MALONEY, L.T. (1999). Physics-based approaches to modeling surface color perception. In *Color Vision: From Genes to Perception*, eds. MALONEY, L.T., GEGENFURTNER, K.R. & SHARPE, L.T., pp. 387–422. Cambridge, UK: Cambridge University Press.
- MCCOLLOUGH, C. (1965). Color adaptation of edge-detectors in the human visual system. *Science* **149**, 1115–1116.

- MOLLON, J.D. & JORDAN, G. (1997). On the nature of unique hues. In *John Dalton's Colour Vision Legacy*, eds. DICKINSON, C., MURRAY, I. & CARDEN, D., pp. 381–392. London, UK: Taylor & Francis.
- NASCIMENTO, S.M.C., FERREIRA, F.P. & FOSTER, D.H. (2002). Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A* **19**, 1484–1490.
- NEITZ, J., CARROLL, J., YAMAUCHI, Y., NEITZ, M. & WILLIAMS, D.R. (2002). Color perception is mediated by a plastic neural mechanism that remains adjustable in adults. *Neuron* **35**, 783–792.
- O'REGAN, J.K. & NOË, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences* **24**, 939–1011.
- PARKKINEN, J.P.S., HALLIKAINEN, J. & JAASKELAINEN, T. (1989). Characteristic spectra of Munsell colors. *Journal of the Optical Society of America A* **6**, 318–322.
- REGIER, T., KAY, P. & COOK, R.S. (2005). Focal colors are universal after all. *Proceedings of the National Academy of Sciences of the U.S.A.* **102**, 8386–8391.
- ROMERO, J., GARCIA-BELTRAN, A. & HERNANDEZ-ANDRES, J. (1997). Linear bases for representation of natural and artificial illuminants. *Journal of the Optical Society of America A* **14**, 1007–1014.
- RUDERMAN, D.R., CRONIN, T.W. & CHIAO, C.-C. (1998). Statistics of cone responses to natural images: Implications for visual coding. *Journal of the Optical Society of America A* **15**, 2036–2045.
- STILES, W.S. & BURCH, J.M. (1959). NPL colour-matching investigation: Final report. *Optica Acta* **6**, 1–26.
- STOCKMAN, A. & SHARPE, L.T. (2000). Spectral sensitivities of the middle- and long-wavelength sensitive cones derived from measurements in observers of known genotype. *Vision Research* **40**, 1711–1737.
- SUGITA, Y. (2004). Experience in early infancy is indispensable for color perception. *Current Biology* **14**, 1267–1271.
- VALBERG, A. (2001). Unique hues: An old problem for a new generation. *Vision Research* **41**, 1645–1657.
- WANDELL, B. (1997). *Foundations of Vision*. Sunderland, Massachusetts: Sinauer Associates.
- WEBSTER, M.A., MIYAHARA, E., MALKOC, G. & RAKER, V.E. (2000). Variations in normal color vision. I. Cone-opponent axes. *Journal of the Optical Society of America A* **17**, 1535–1545.
- WESTLAND, S., SHAW, A.J. & OWENS, H.C. (2000). Colour statistics of natural and man-made surfaces. *Sensor Review* **20**, 50–55.
- ZEKI, S. (2003). The disunity of consciousness. *Trends in Cognitive Science* **7**, 214–218.

Appendix: Real and complex eigenvalues

We recall that the eigenvalues of a linear operator A on a vector space V are the roots (possibly complex) of the characteristic polynomial $P(x) = \det(A - x \cdot Id)$. In the case where all roots have multiplicity 1, then there is a basis of V made of corresponding eigenvectors, that is, non-null vectors u_i such that $Au_i = x_i \cdot u_i$ for some complex values x_i . Only a set of operators with null measure do not have roots with multiplicity 1: hence, with probability 1, A can be decomposed in terms of complex eigenvalues and eigenvectors. In the case where V is three dimensional and A is real, then A has one real root and the two other roots are either real, or conjugate: $\rho e^{i\theta}$ and $\rho e^{-i\theta}$. Recalling that from a norm $\|\cdot\|$ on V , it is possible to define a norm on linear operators by $\|A\| = \max\{\|Au\|, \|u\| = 1\}$, it can be shown that for all possible norms on V , the distance $\|A - \bar{A}\|$ between an operator A with complex eigenvalues and the closest operator \bar{A} with real eigenvalues is no greater than $\rho \sin(|\theta|/2)$. Hence, if this value is small (compared to 1: it is a quantity without units), then A is “approximately” diagonal in the real field, that is, there is a basis of V such that the action of A on any element u is well approximated by the mere scaling of the coordinates of u in this basis.

Supplementary Information

Singularity index

We required an index characterising surfaces that essentially reflect either one or two components of the available information. We considered the two ratios $\beta_1 = r_1/r_2$ and $\beta_2 = r_2/r_3$, where (r_1, r_2, r_3) are the attenuation values sorted in decreasing order (Fig. 4). Maxima of β_1 reveal surfaces that reflect one of the components much more significantly than the two others, maxima of β_2 reveal surfaces that reflect two of the components much more significantly than the third.

To gather these two kinds of singularity in one index, Figure 1 plots $\max(\beta_1/\beta_1^{\max}, \beta_2/\beta_2^{\max})$, where β_1^{\max} and β_2^{\max} are the maximum values taken by each ratio for an extensive set of surfaces. While providing a unique plot to compare with data from the World Colour Survey, this procedure hides interesting feature of Figure 4. Indeed, it appears that cyan and magenta surfaces also correspond to singular reflecting properties (Fig. 4b) although this is much less noticeable than for red, yellow, green and blue. Further, focal blue appears to be correlated with the existence of a weak singularity of the second type ($r_2 \gg r_3$, Fig. 4b) close to a major singularity of the first type ($r_1 \gg r_2$, Fig. 4a). This unclear situation only happens for blue, and this could be part of an explanation for why blue is the last colour to be singled out in the main line of evolution of languages¹.

To transpose our singularity index from surfaces to LMS tristimuli, we assumed an illuminant daylight at 6500 K, that is D65. As the ratios of reflection coefficients are by construction invariant by scalar multiplication of the reflection matrix, and because therefore the transposed index is expected to be invariant along straight lines going through the origin, we need only to plot the index on some plane cutting the LMS cone. It must of course be understood that our results do not depend on the choice of such a plane, used only to *represent* the data in a convenient way. We chose the plane of equation $L+M+S=1$ merely for convenience, because it yielded the more readable plot. Figure 2a of the article has been slightly retouched to correct visual artefacts at the boundaries of the surface caused by the square grid sampling method of Matlab, and Figure 5 here shows the unmodified original surface. Crests of the surface were fitted visually through the inflexion points of the iso-height contours for red, blue and green. The crest for yellow was fitted visually through the associated local maximum of the index, as iso-height contours do not have clear inflexion points in this area. For clarity, Figure 2a and Figure 5 display the transposed index scaled by a power law. This of course does not modify level sets nor maxima and crests of the index.

We chose the described singularity index for its simplicity, but other indexes could have been used. Indeed, a distinction natural to draw from the

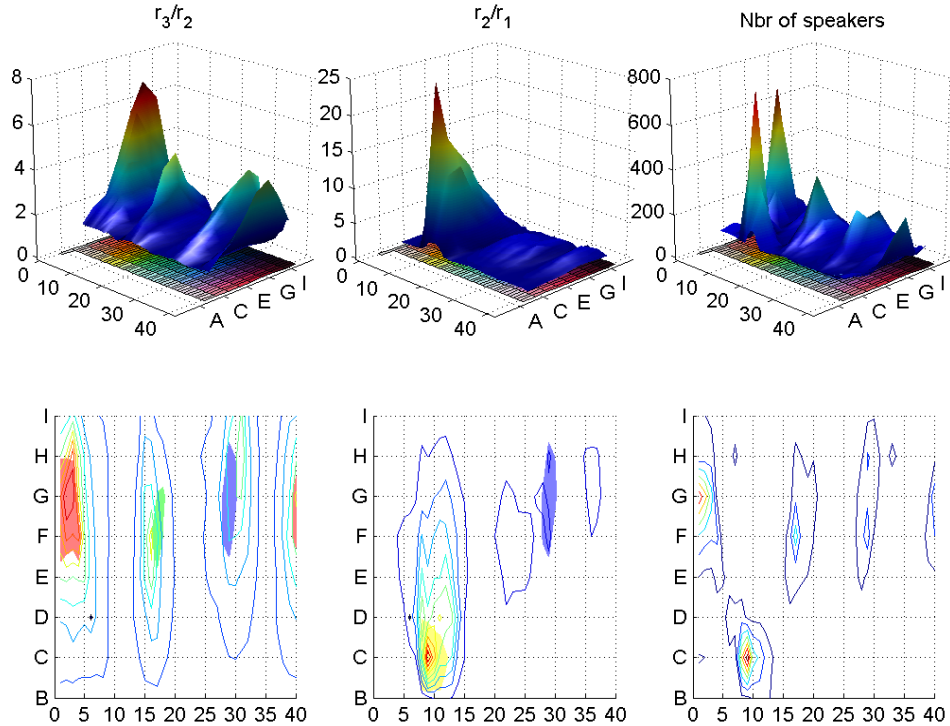


Figure 4 The two kinds of singularity in surface reflecting properties, and colour terms across languages. Left panels show the ratio of the first and second reflecting coefficients of surfaces, middle panels show the ratio of the second and third reflecting coefficients, right panels show data from the Berlin and Kay World Colour Survey¹¹. Focal red, yellow and green correspond to clear distinct maxima of either one or other kind of singular reflecting property of surfaces. Focal blue corresponds to the conjunction of two weaker singularities of each kind.

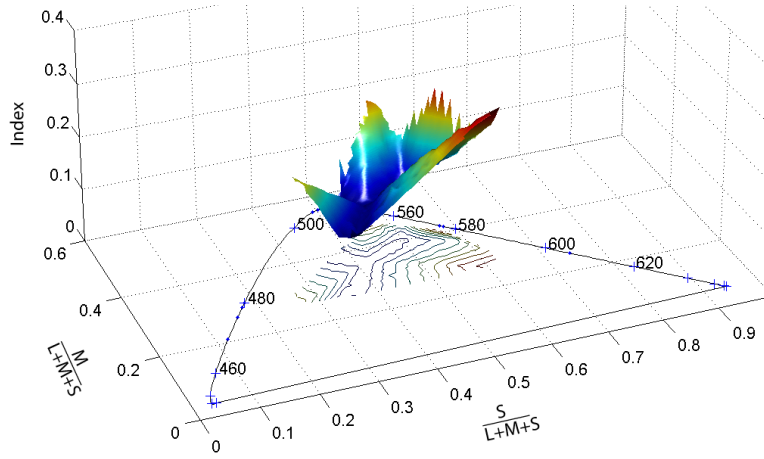


Figure 5 Singularity index calculated for photon catches L, M and S of human photopigments. This is Figure 1, before correction of visual artefacts at the boundaries of the surface caused by the square grid sampling method of Matlab.

mathematical viewpoint is the distinction between cases when the reflection matrix has one eigenspace of dimension three, or one eigenspace of dimension two and one of dimension one, or three eigenspaces of dimension one. The general case is the last one, and the surfaces that we expect to be considered as perceptually singular are therefore the ones falling in the two first cases. The case with one eigenspace of dimension three corresponds to achromatic surfaces (meaning that illuminants will be merely attenuated by the surface with no change of chromaticity). The case with one eigenspace of dimension two and one of dimension one corresponds to the case studied in the paper. Because the dimension of the eigenspaces reflects the equality of some eigenvalues, one would be interested in an index having value 1 when two eigenvalues are equal and a value close to 0 otherwise. Such an index is given by

$$\text{Ind} = | (r_1 - 2r_2 + r_3) / (r_1 - r_3) |$$

The predictions reported in Table 1 were derived from this ‘closer examination of the singularities of surface reflecting properties’, shown in Figure 6.

Separately attenuated components

The spectral composition of light is classically represented by elements of the real linear space of

square-integrable functions; the information available about light can be represented as an n -dimensional real vector specifying for instance the photon absorption count for each of the n photopigment types that an organism possesses. Let u be the linear mapping between illuminants and the information available via a given set of photopigments; for a given surface of reflectance $S(\lambda)$, let v^S be the linear mapping between illuminants and the information available about the lights reflected by that surface:

$$u_i(E) = \int R_i(\lambda) E(\lambda) d\lambda \quad \text{and} \quad v_i^S(E) = \int R_i(\lambda) S(\lambda) E(\lambda) d\lambda$$

where R_i are photopigment sensitivities. What we call *non-interacting components* C_i^S of the information available about the illuminants are components that satisfy for any illuminant $E(\lambda)$:

$$u(E) = \sum_i \alpha_i C_i^S \Rightarrow v^S(E) = \sum_i r_i \alpha_i C_i^S.$$

The idea is that such components are the biological organism's analogue of the physicist's decomposition of light into monochromatic components: As seen from behind the organism's receptor apparatus, such components have the property that they do not interact and are simply independently attenuated by the reflecting surface.

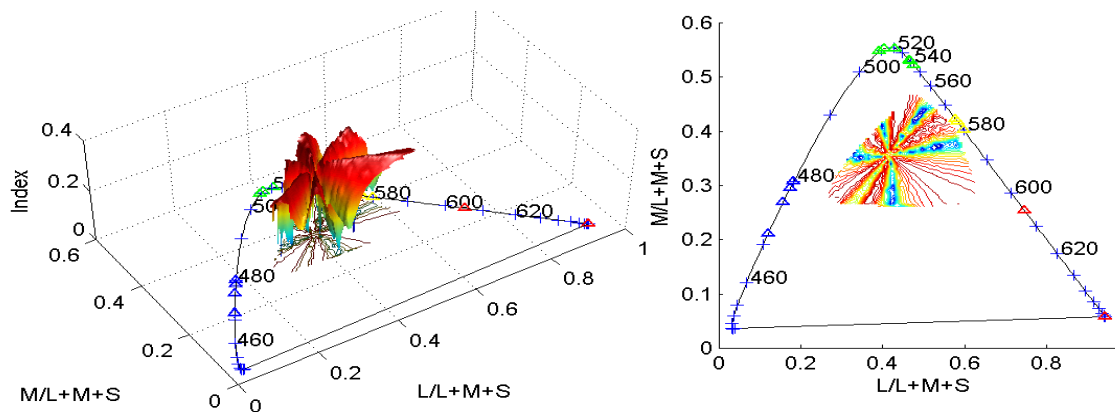


Figure 6 A closer look on singular reflecting properties: a variant of the singularity index, that pinpoints tristimuli associated with surfaces that have two reflection coefficients equal (but not necessarily with a large difference between the two different coefficients).

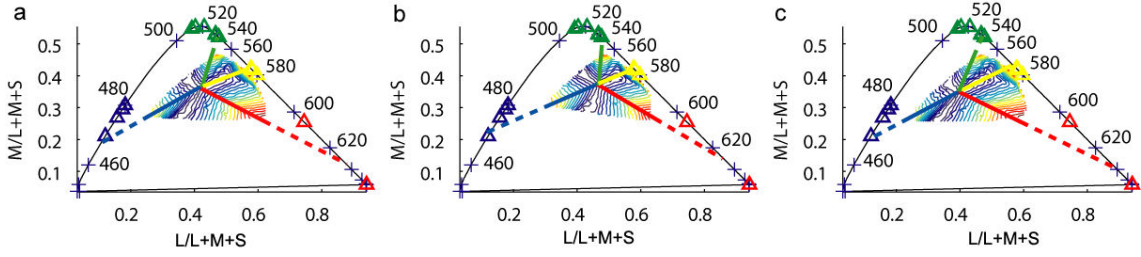


Figure 7 Dependence of the result on the illuminant assumed to transpose the singularity index from surfaces to photopigment catches. **a**, natural daylight of temperature 5500 K. **b**, natural daylight of temperature 6500 K. **c**, natural daylight of temperature 7500 K.

For a given set of illuminants $\{E_i(\lambda)\}$, consider the set of couples $\{(u(E_i), v^s(E_i))\} \subset \mathbb{R}^n \times \mathbb{R}^n$. If there is an n -dimensional subspace that well approximates this set of couples, it can be interpreted as the graph of a linear mapping a^s between the available information $u(E)$ and $v^s(E)$, so that $v^s(E) \approx a^s(u(E))$ (this is nothing but a multilinear regression). In this case

$$u(E) = \sum_i \alpha_i C_i^s \Rightarrow v^s(E) = a^s(u(E)) = \sum_i \alpha_i a^s(C_i^s)$$

and non-interacting components must satisfy $a^s(C_i^s) = r_i C_i^s$: they must be eigenvectors of a^s with eigenvalues r_i .

The 400nm-700nm band of daylight spectra are known to be well approximated by weighted sums of only a few square-integrable functions²⁻⁵. In these conditions, and as concerns humans, there is thus a fortiori a three dimensional subspace that provides an excellent approximation for the set of couples $\{(u(E), v^s(E))\}$ obtained when considering daylight illuminants. Note however that low dimensionality of the illuminant space is not necessary to have low dimensionality of the set of couples: it would be quite possible for the illuminants to require a large number of basis functions, but for the couples $\{(u(E), v^s(E))\}$ still to occupy a low dimensional space. This is because the information available about an illuminant provides only poor information about its spectrum.

In any case the incidental three dimensionality of illuminants gives a straightforward way to compute the eigenvalues of a^s . Noting $\{E_0, E_1, E_2\}$ a basis for the illuminants and considering the $\{L, M, S\}$ basis for the information available, then we find 3 by 3 matrices U and V which are representations of the linear mappings u and v :

$$U_{ij} = \int R_i(\lambda) E_j(\lambda) d\lambda \quad \text{and} \quad V_{ij} = \int R_i(\lambda) S(\lambda) E_j(\lambda) d\lambda$$

where R_L , R_M and R_S are the photopigment sensitivity curves of the L, M, and S photopigments. Then the matrix representation of a^s in the $\{L, M, S\}$ basis is given by $V U^{-1}$, which allows the eigenvectors and eigenvalues of a^s to be computed.

Qualitative stability of the results

Here we assess the qualitative stability of our results with respect to the illuminant that we considered in the

inference scheme and the dataset of surfaces and daylights considered. Before that, it may however be useful to note that the results presented here obviously demand to be refined on two grounds. First, the inference scheme considered here is very naïve: the nervous system might use different hypotheses about the illuminant for different sensory inputs. Second, all the results presented here rely on the reflection coefficients alone, without any consideration for the eigenvectors, that can be different from surface to surface. This last point is obviously very important, and will require serious further examination.

We could have chosen a different illuminant to transform the singularity index of surfaces into the singularity index for tristimuli, for instance daylights of various temperature such as D55, D65 and D75. This choice does not however modify significantly the predicted monochromatic unique hues, though of course it modifies the position of white and thus the equation of the four half planes (Fig. 7).

The index was plotted using a dataset of 1600 Munsell chips because this dataset contains coloured surfaces spanning a significant area within the LMS cone. However it is known that Munsell chips do not provide a representative set for natural reflectances. Figure 8 shows the index transposed from a larger dataset of surfaces including 1077 natural reflectances of flowers, leaves, wood, plants, etc: 218 reflectances from the University of Joensuu⁶, 118 reflectances from the University of Trieste⁷, 404 reflectances from S. Westland⁸ and 337 from the Krinov dataset⁹.

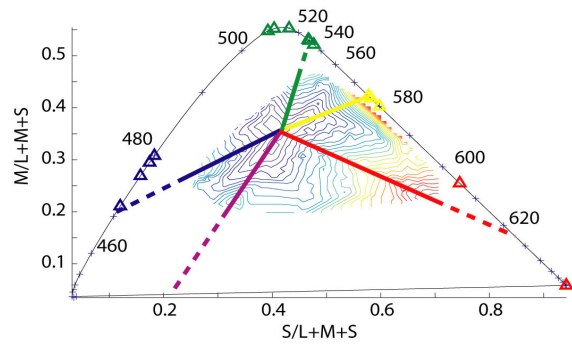


Figure 8 Singularity index transposed from surfaces to photopigment catches, based on a dataset of surfaces including 1600 Munsell chips and 1077 natural reflectances of flowers, leaves, beans, etc. The four maxima observed when using Munsell chips remain unchanged, while an additional maximum appears that corresponds to magenta.

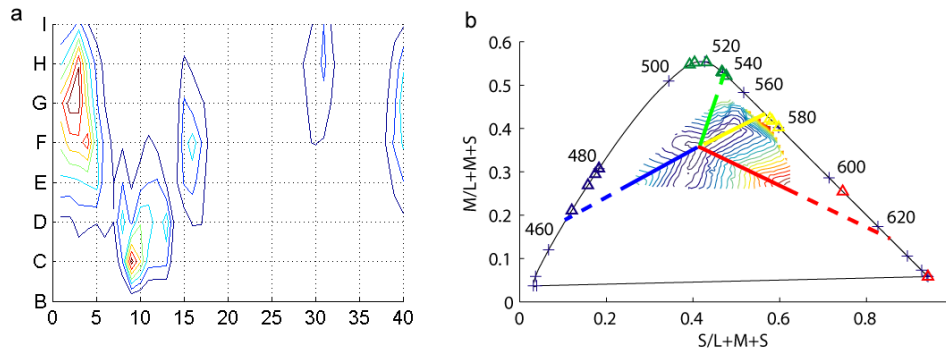


Figure 9 Predictions of focal colours and unique hues, using only the set of illuminants spanned by the basis from Judd & al.³. **a**, focal colours. **b**, unique hues.

The figures presented in the article all rely on a set of illuminants compiling the measurements of Judd & al.⁴, Romero & al.⁵, and Chiao & al.³ Figure 9 shows the predictions.

Hue cancellation

While our approach can be seen to correctly predict that yellow-blue equilibrium (i.e. red-green unique hue locus) should not be closed under addition while red-green equilibrium (i.e. Yellow-blue unique hue locus) should, we made some simplifications that amount to predict closure under scalar multiplication, which is known not to hold for yellow-blue equilibrium¹⁰. This was done when saying that the transposed index is *expected* to be invariant along straight lines going through the origin because the ratios of reflection coefficients are by construction invariant by scalar multiplication of the reflection matrix. This was taken as a simplification to present our approach and derive our predictions of psychophysical experiments, but it is actually expected to be only roughly true. Because the dimension of the set of natural reflectances is known to be of dimension greater than 3, several reflectances (hence different triplets of attenuation values) will necessarily be associated with individual tristimuli. If one considers that the nervous system has learnt to expect from a given tristimulus the reflectance that is most probably associated with it, it may be that these most probable reflectances are not closed by scalar multiplication (e.g. they lie in a non-linear manifold). As a result the index value associated with α .LMS may be different from the index value of α times the reflectance associated with LMS.

Matlab files

All the Matlab files and data used to compute the figures of the article and of this Supplementary Information can be downloaded at <http://nivea.psycho.univ-paris5.fr/~philipona/repository/matlabFiles.zip>. Set the path for Matlab to access all the subdirectories of the unzipped files, and type “plotAllFigures;”.

Bibliography

1. Kay, P. & Maffi, L. Color Appearance and the Emergence and Evolution of Basic Color Lexicons. *American Anthropologist* **101**, 743-760 (1999)
2. Maloney, L. T. Physics-based approaches to modeling surface color perception. In *Color Vision: From Genes to Perception* (Cambridge Univ. Press, Cambridge U.K., 1999)
3. Chiao, C.-C. & Cronin, T.W. & Osorio, D. Color signals in natural scenes: characteristics of reflectance spectra and effects of natural illuminants. *J. Opt. Soc. Am. A* **17**, 218-224 (2000)
4. Judd, D. B. & MacAdam, D. L. & Wyszecki, G. Spectral distribution of typical daylight as a function of correlated color temperature. *J. Opt. Soc. Am.* **54**, 1031-1040 (1964).
5. Romero, J. & Garcia-Beltran, A. & Hernandez-Andres, J. Linear bases for representation of natural and artificial illuminants. *J. Opt. Soc. Am. A* **14**, 1007-1014 (1997)
6. Parkkinen, J. & Jaaskelainen, T. & Kuittinen, M. Spectral representation of color images. IEEE 9th International Conference on Pattern Recognition, Rome, Italy, **2**, 933-935 (1988)
7. http://mitglied.lycos.de/whiteness/Internet_sites/internet_sites.html
8. Westland, S. & Shaw, A. J. & Owens, H. C. Colour statistics of natural and man-made surfaces. *Sensor Review* **20**, 50-55 (2000)
9. E. L. Krinov, Spectral Reflectance Properties of Natural Formations. *National Research Council of Canada* (1947)
10. Ejima, Y. & Takahashi, S. Interaction between short- and longer-wavelength cones in hue cancellation codes: nonlinearities of hue cancellation as a function of stimulus intensity. *Vision Res.* **25**, 1911-22 (1985)

Annexe A

L'étude des couleurs

L'étude de la perception des couleurs a une longue histoire : Platon avait déjà une théorie sur la question. Mais c'est seulement au *XVII^{ème}* siècle que commence à s'articuler une argumentation appuyée sur des faits expérimentaux non triviaux. La difficulté de cette argumentation est qu'il fallait dans le même temps comprendre la physique de la lumière, et la perception des couleurs : comprendre que l'observation des résultats d'expériences était déterminé à la fois par des faits physiques et des contraintes propres à notre système visuel.

L'aperçu historique suivant de l'étude des couleurs est, pour les sections [A.1](#) et [A.2](#), fortement inspiré de l'excellente revue de [Mollon \(2003\)](#), complétée par diverses sources, en particulier [Grassmann \(1854\)](#); [Judd \(1966\)](#); [Kaiser and Boynton \(1996\)](#). L'intérêt de cet historique, outre celui de donner des éléments de référence, est de montrer à quel point la science des couleurs, tout en progressant dans l'accumulation de données psychophysiques et dans les liens trouvés avec la physique du monde et la physiologie de notre système sensoriel, n'est jamais qu'une suite de théories vis-à-vis desquelles il faut rester critique : des théories dépendantes de l'évolution de paradigmes conceptuels généraux, dont certains nous semblent, du point de vue actuel, incompréhensibles ou absurdes.

A.1 Les balbutiements d'une théorie de la couleur

A.1.1 La découverte du spectre

En 1671 est publié un article de Newton sur la décomposition de la lumière. L'arc-en-ciel obtenu à partir d'un faisceau de lumière blanche et d'un prisme était déjà célèbre à l'époque, mais Newton décrit dans son article comment, en plaçant derrière le prisme deux plaques percées d'un trou, il a obtenu un faisceau lumineux très particulier (Fig. A.1(a)). Ce faisceau ne donne plus lieu à un spectre de dispersion s'il est envoyé sur un second prisme, et la teinte du halo lumineux obtenu sur diverses surfaces colorées est toujours la même. Aucune des opérations qu'il imagine ne permettent de changer cette teinte.

Newton conclut de ces expériences que la lumière blanche est composée de lumières élémentaires, qui sont séparées par le prisme. Il ajoute que la couleur est une disposition de ces différentes lumières élémentaires à évoquer en nous une certaine impression, et argumente que la couleur d'un objet est simplement la conséquence de ce que cet objet renvoie plus ou moins les différentes lumières élémentaires. Ce point de vue est en opposition avec toutes les théories précédentes des couleurs, qui tenaient la couleur pour une caractéristique des objets.

Sur la base de son expérience, Newton distingue deux sortes de couleurs : les couleurs prismatiques observables dans le spectre de dispersion, qui ne peuvent plus être altérées, et les couleurs composées, comme le blanc. En 1704, Newton introduit l'ancêtre des diagrammes chromatiques, destiné à prédire la couleur de la superposition de différentes lumières (Fig. A.1(b)). Newton donne à ce diagramme la forme d'un cercle, parce que les couleurs prismatiques allant du rouge au violet elles semblent se rejoindre aux deux extrémités du spectre. Newton catégorise l'ensemble des couleurs prismatiques en 7 classes : rouge, orange, jaune, vert, bleu, indigo, violet.

Newton affirme à propos du blanc :

There is no one sort of Rays which alone can exhibit this. Tis ever compounded, and to its composition are requisite all the aforesaid [prismatic] colours.

Or Huygens critique rapidement la seconde affirmation : dès 1673, il suggère que le blanc peut être obtenu par superposition de juste deux lumières prismatiques.

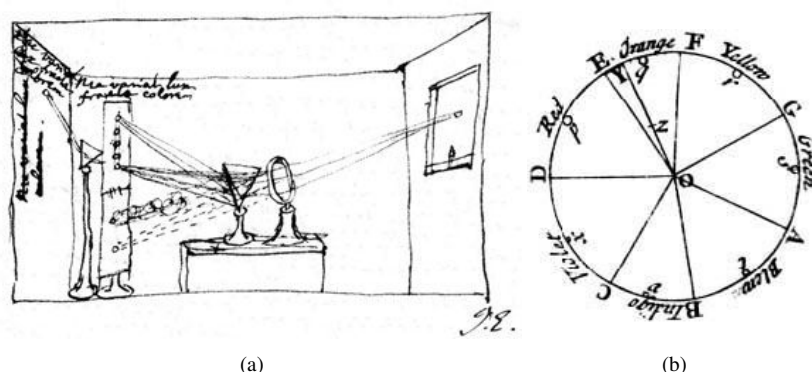


FIG. A.1 – L'expérience de Newton, et le cercle des couleurs, 1704.

A.1.2 La notion de couleurs primaires

C'est l'imprimerie qui fera avancer notre compréhension de ce qui est aujourd'hui désigné sous le terme de *trichromaticité*. En 1719, Jacques Christophe Le Blon dépose le premier brevet sur l'impression couleur à partir de trois plaques encrées respectivement de jaune, de rouge et de bleu. Dans un traité de 1708, anonyme mais sans doute dû à Le Blon, l'auteur détaille sa conception des couleurs :

Il n'y a proprement que trois Couleurs Primitives, lesquelles ne peuvent pas êtres composées par d'autres couleurs, mais dont toutes les autres peuvent être composées. Ces trois couleurs sont le Jaune, le Rouge et le Bleu, car pour le Blanc et le Noir, ce ne sont pas proprement des Couleurs, le Blanc n'étant autre chose que la représentation de la lumière, et le Noir la privation de cette lumière

Pour les besoins de l'imprimerie, on se met à établir des diagrammes spécifiant les proportions d'encre jaune, rouge et bleu à superposer pour obtenir les différentes couleurs à reproduire. Tobias Mayer en 1758, puis Johann Heinrich Lambert en 1770 (Fig. A.2), produisent des diagrammes en forme de pyramide où les coordonnées spatiales du diagrammes spécifient les proportions d'encre à utiliser.

L'idée que c'est en raison des particularités du système visuel, et pas de la physique de la lumière elle-même, que le mélange de seulement trois couleurs primaires permettait de reproduire toutes les couleurs n'a émergé que très progressivement. Dans un article de l'académie impériale de St Petersburg, Mikhail Vasil'evich Lomonosov suggère en 1756 que la lumière consiste en la vibration de trois types de particules, éga-

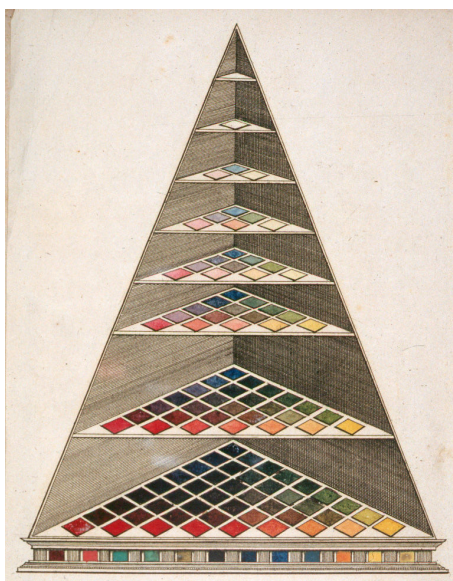


FIG. A.2 – La pyramide de Lambert pour le mélange des pigments.

lement présentes dans l'oeil. En 1777, George Palmer suggère qu'il existe trois types de lumières, et trois types correspondants de "molécules" ou "membranes" dans la rétine. Il est sans doute l'un des premiers à proposer comme explication du daltonisme un problème avec l'un de ces trois récepteurs hypothétiques. Mais c'est seulement trente ans plus tard que Thomas Young articulera une théorie ondulatoire de la lumière avec la présence de trois types de "résonateurs" rétinien, sensibles sur une large plage de longueur d'onde mais ayant des pics à différents endroits du spectre. Young affirmera d'abord que les trois couleurs "principales" auxquelles répondent ces résonateurs sont le jaune, le rouge et le bleu, puis il suggèrera cinq ans plus tard qu'il s'agit du rouge, du vert et du violet.

A.1.3 Le rôle du contexte dans la perception des couleurs

En 1789, Gaspard Monge fait à l'Académie des Sciences de Paris une démonstration qui va mettre en évidence la complexité insoupçonnée de notre perception des couleurs. Il demande aux académiciens d'observer une scène au travers d'un filtre rouge, pour leur faire constater de visu que les objets qui apparaissaient rouge à la lumière du jour semblent presque blanc derrière le filtre. Pour expliquer ce fait, Monge suggère que notre perception des couleurs ne dépend pas d'une manière triviale de la lumière

qui entre dans nos yeux. Selon lui, nous interprétons cette lumière en fonction de ce que nous pensons être l'illuminant de la scène : si nous pensons que l'illuminant est rougeâtre, alors nous identifions comme blanc ce qui sous la lumière du jour nous aurait semblé rouge :

Ainsi les jugements que nous portons sur les couleurs des objets ne paraissent pas dépendre uniquement de la nature absolue des rayons de lumière qui en font la peinture sur la rétine ; ils peuvent être modifiés par les circonstances, et il est probable que nous sommes déterminés plutôt par la relation de quelques-unes des affections des rayons de lumière, que par les affections elles-mêmes, considérées d'une manière absolue.

Le fait que le jugement de couleur que nous portons sur les objets est largement indépendant de l'illuminant sous lequel ces objets sont observés est aujourd'hui désigné par le terme de *constance des couleurs*. C'est une caractéristique fondamentale de la perception des couleurs, dont les premières observations remontent à la fin du *XVII^{ème}* siècle : en 1694, De la Hire écrit que deux objets identiques observés simultanément sous différents illuminants ont une apparence différente, mais que l'on est incapable de noter cette différence si on les observe isolément. Ce qu'il est intéressant de noter, c'est que Monge exprime dans l'extrait précédent l'opinion que les couleurs sont un jugement que nous portons sur les objets, jugement réalisé par l'intermédiaire de la lumière entrant dans l'œil. Cette position revient ainsi sur le mouvement conceptuel opéré par Newton, qui défendait l'idée que la couleur est une qualité de la lumière elle-même.

Monge se penche également sur une expérience décrite en 1672 par Otto von Guericke de Magdeburg, qui va dans le sens de sa théorie selon laquelle le contexte chromatique est crucial pour déterminer l'apparence d'un stimulus lumineux. Il s'agit du phénomène des "ombres colorées" : lorsqu'un objet intercepte la lumière provenant d'une lumière rouge et d'une lumière blanche, on observe sur un écran placé derrière l'objet deux ombres de couleur rouge et verte sur un fond rose. Alors que l'ombre rouge s'explique naturellement du fait que l'objet bloque la lumière blanche, ne laissant que la lumière rouge arriver à cet endroit de l'écran, l'ombre verte est surprenante. En réalité, si l'on regarde uniquement cette ombre au travers d'une plaque noire trouée, elle apparaît alors comme blanche, ce qui est la couleur à laquelle on s'attend par le raisonnement précédent. La perception d'une ombre verte est donc la conséquence des

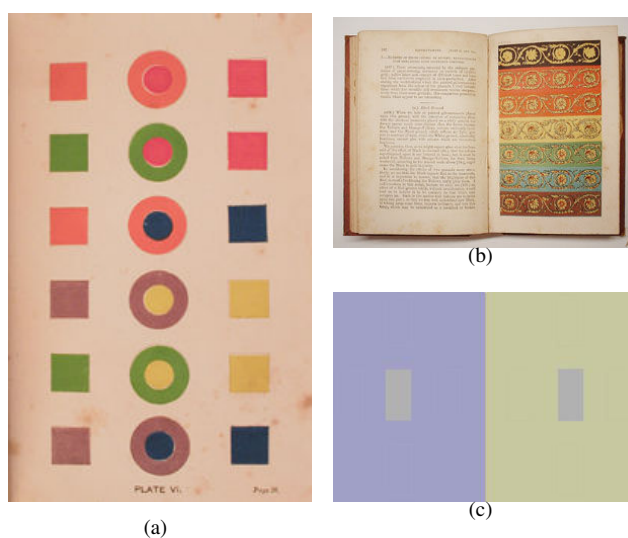


FIG. A.3 – Effet du contraste sur la perception des couleurs. a) Chevreul, 1861, b) exemple contemporain : les deux rectangles centraux ont la même composition spectrale.

caractéristiques de notre système visuel.

Pour des questions manufacturières et artistiques, on se met alors à étudier les règles décrivant l'influence du contexte chromatique sur la couleur perçue. Le chimiste Michel-Eugène Chevreul, nommé par Louis XVIII directeur de la manufacture royale des Gobelins, publie en 1839 un livre en trois volumes sur "la loi du contraste simultané des couleurs". Il y explique entre autre que quand une aire colorée est entourée d'un fond d'une teinte différente, la couleur perçue est biaisée significativement dans la direction de la teinte située sur le cercle de Newton à l'opposé de la couleur du fond (Fig. A.1.3).

L'époque est également marquée par l'opposition véhémente de Johann Wolfgang von Goethe à la théorie Newtonienne des couleurs : en 1810 il publie sa propre théorie, qui restera cependant marginale. En 1824, Jan Evangelista Purkinje identifie un effet curieux qui gardera son nom, l'effet Purkinje : lorsque l'oeil passe des conditions d'adaptation à une forte illumination à celle d'une faible illumination, un rouge perçu comme plus lumineux qu'un vert va finalement devenir moins lumineux que ce dernier.

A.2 Trichromaticité et opposition

A.2.1 Le mélange additif des couleurs

A partir du milieu du *XIX^{ème}* siècle, la compréhension de la perception des couleurs entre dans une nouvelle ère, marquée par la mise en oeuvre de paradigmes expérimentaux quantitatifs. C'est à Hermann von Helmholtz qu'on doit en 1852 d'avoir tout d'abord clarifié la distinction entre mélange soustractif et additif des couleurs, concernant le mélange des pigments d'une part et la superposition des lumières d'autre part. Il démontre cette différence en utilisant astucieusement deux peintures, bleue et jaune, et un disque. Il peint le centre du disque avec un mélange des deux peintures, et la circonférence avec des secteurs alternés de chacune des peintures, puis il montre le disque en train de tourner. Le centre apparaît vert et la circonférence gris clair - dans un cas l'oeil reçoit la lumière réfléchie par le mélange des peintures, et dans l'autre cas l'oeil intègre alternativement les lumières réfléchies individuellement par chaque peinture, ce qui produit une stimulation équivalente à celle produite par la superposition de ces lumières.

Helmholz s'est également penché sur la question, soulevée par Newton, de l'obtention du blanc à partir de la superposition de seulement deux lumières élémentaires. En 1852 il décrit dans un article comment, en croisant deux spectres de dispersion de diverses luminosités, il a observé toutes les combinaisons possibles de lumières monochromatiques. Il n'a trouvé par cette méthode que deux lumières monochromatiques, jaune et bleu, dont la superposition donne une lumière blanche - ce que l'on appelle des couleurs complémentaires. Ce résultat appellera l'année suivante une brillante objection de Hermann Grassmann, basée sur une considération topologique.

Grassmann fait la remarque que, lorsqu'on ajoute à une lumière prismatique une autre lumière prismatique d'intensité croissante, on obtient une série continue de teintes joignant les teintes des deux lumières, série qui ne peut être que de trois types : soit les teintes évoluent dans l'ordre rouge-jaune-vert-bleu (qu'il appelle le sens "direct"), soit dans l'ordre inverse (sens "indirect"), soit en passant par l'achromatique. Bien sûr, la série joignant une teinte spectrale à celle qui est juste après elle dans le sens direct est une série directe, et la série joignant une teinte spectrale à celle qui est juste avant elle est une série indirecte. Comme Grassmann tient l'ensemble des teintes spectrales

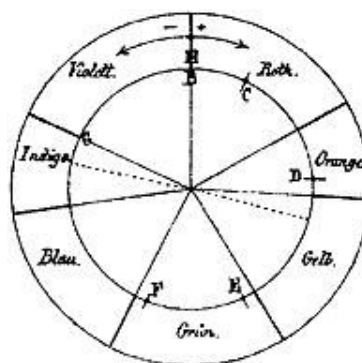


FIG. A.4 – Le cercle des couleurs de Grassmann

pour être une courbe fermée¹, il en conclut qu'il y a nécessairement une certaine teinte spectrale pour laquelle la série passe par l'achromatique (Fig. A.2.1). Helmholtz reconsidère alors ses expériences, et trouve en 1855 que tout un ensemble de couleurs monochromatiques ont effectivement des couleurs complémentaires. C'est à Grassmann que l'on doit par ailleurs d'avoir clairement formulé que si deux lumières ont une couleur identique, alors elles auront encore une couleur identique après qu'il leur soit superposé une même lumière – ce qui démontre qu'un processus linéaire par rapport aux spectres lumineux est au cœur de nos jugements sur les couleurs.

A.2.2 L'égalisation de la couleurs des lumières

La même année, James Clerk Maxwell réalise une expérience sur le mélange additif des lumières, inspirée par le disque de Helmholtz. Il utilise un disque composé à la périphérie de trois secteurs colorés d'aires variables, et au centre de deux secteurs noirs et blancs également d'aires variables. Lorsqu'il tourne, le disque a l'apparence d'un anneau coloré autour d'un centre gris, la couleur de l'anneau et le gris du centre pouvant être ajustés en modifiant l'aires des secteurs. En cherchant les manières d'ajuster les aires périphériques pour que la couleur de l'anneau soit indiscernable de celle du centre, Maxwell obtient une équation de la forme : $.37V + .27U + .36EG = .28SW + .72Bk$, indiquant que si la périphérie du disque est constituée d'un secteur de 37 radians de vermillon (V), 27 radians d'ultramarine (U) et 36 radians de vert émeraude (EG), cela est indiscernable d'un centre constitué de 28 radians de blanc et 72 radians de noir. En réalisant de nouveau l'expérience avec une couleur test (T) à la place d'une des

¹Ce qui n'est dans la réalité pas exact.

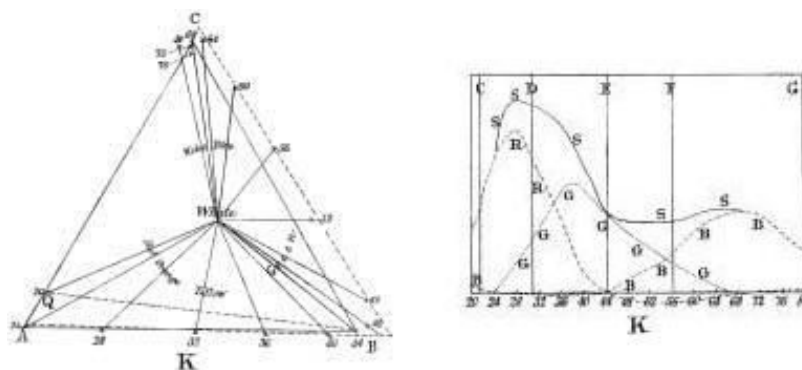


FIG. A.5 – Les CMFs mesurées par Maxwell.

couleurs de base, on obtient une seconde équation qui permet d'éliminer la variable SW , et d'exprimer T en fonction de quatre coefficients pour U , V , EG et Bk (ce qui est bien tri-dimensionnel, la somme des coefficients des couleurs primaires étant par construction égale à 1).

En 1860, Maxwell construit un nouveau dispositif qui permet de quantifier la contrainte trichromatique du système visuel d'une manière plus directe. Son dispositif lui permet de réaliser l'égalisation d'apparence entre une lumière quelconque et la superposition de trois lumières nommées primaires, mesurant ainsi pour la première fois ce qui sera nommé par la suite un ensemble de "Color Matching Functions" (CMFs). L'idée est de mesurer l'intensité des primaires qui permet d'égaliser l'apparence de chaque lumière monochromatique, ce qui permet ensuite, à partir des lois de Grassmann, de prédire l'intensité nécessaire pour égaliser l'apparence de n'importe quelle lumière de composition spectrale donnée, à partir de n'importe quelles primaires. L'intensité des trois primaires donne un système de coordonnées tri-dimensionnelles pour représenter l'ensemble des couleurs, une représentation qui prend la forme moderne d'un diagramme chromatique (Fig. A.5).

Arthur König, un collègue de Helmholtz, affine les mesures de Maxwell, et les complète par la mesure de CMFs pour deux types de daltoniens, protanopes et deutanopes. Ces données lui permettront en 1886 de produire les premières estimations psychophysiques des courbes de sensibilités des récepteurs considérés dans la théorie de Young, théorie reprise et popularisée par Helmholtz. König montre par ailleurs en 1894 que la courbe de visibilité scotopique – la courbe décrivant l'ajustement d'intensité nécessaire, lorsqu'on change la longueur d'onde d'une lumière monochromatique,

pour maintenir une apparence identique des deux lumières en vision de très faible niveau d'intensité – est compatible avec les mesures réalisées par Kühne en 1878 sur l'absorption de la rhodopsine trouvée dans certaines cellules rétiniennes.

A.2.3 La théorie de Hering

Malgré ses succès, la théorie de Young-Helmholz conserve quelques détracteurs. Le physiologiste Ewald Hering soutient que, du point de vue phénoménologique, il y a quatre couleurs simples : rouge, vert, jaune et bleu. Selon lui, on peut voir du rouge et du bleu "dans" le violet, alors que l'on ne peut rien voir d'autre "dans" le rouge que le rouge. Que le jaune soit considéré dans la théorie trichromatique comme une sensation simultanée de rouge et de vert lui paraît inacceptable. Il ajoute que, dans la décomposition phénoménale à laquelle il se réfère, on ne trouve jamais simultanément de rouge et de vert, de jaune et de bleu, et pour cette raison il qualifie ces couleurs d'*opposées*. Hering suggère dans une série d'articles en 1878 que ces "faits phénoménologiques" sont la conséquence de deux mécanismes neuronaux distincts, l'un associé aux sensations rouge/vert et l'autre aux sensations jaune/bleu. Cette théorie sera considérée avec intérêt soixante-dix ans plus tard, au regard de certaines particularités de la connectivité rétinienne.

Hering s'oppose également à Helmholtz sur la question des contrastes simultanés étudiés par Chevreul. William James résume ainsi les deux positions dans son fameux *Principles of Psychology* (James, 1890) :

Helmholtz maintains that the neural process and the corresponding sensation also remain unchanged, but are differently interpreted ; Hering, that the neural process and the sensation are themselves changed, and that the 'interpretation' is the direct conscious correlate of the altered retinal conditions. According to the one, the contrast is psychological in its origin ; according to the other, it is purely physiological.

A.3 L'étude des capacités visuelles

A.3.1 Contrastes successifs et adaptation du système visuel

En addition des effets de contraste simultané déjà évoqués, Helmholtz note l'existence de forts effets de contraste successif : la perception des couleurs dépend des stimuli auxquels l'oeil a été précédemment exposé. Il l'explique par une fatigue de la substance nerveuse réagissant à la lumière. Johannes von Kries et König montrent, dans les années 1890, que ces effets n'affectent cependant pas les égalisations réalisées par les observateurs. En 1905, von Kries propose un modèle quantitatif de cette fatigue en termes d'un gain pour chaque récepteur, spécifiant l'état d'adaptation du système visuel. Selon cette théorie, l'étude de l'adaptation chromatique fournit une nouvelle manière d'estimer la courbe de sensibilité des récepteurs rétiniens. Cette idée sera poussée à son terme pour la première fois en 1934 par Wright, mais la méthode donne des résultats en désaccord partiel avec celle exploitant les données issues d'observateurs daltoniens.

En 1939, Walter Stiles publie le résultat de ses expériences sur les limites de nos capacités de détection chromatique : il a mesuré, pour un ensemble de conditions d'adaptation, l'intensité de chaque lumière monochromatique nécessaire pour qu'un bref flash de cette lumière soit détecté. La modification de nos capacités de détection chromatique en fonction des conditions d'adaptation lui suggère d'abord que la détection s'appuie sur trois mécanismes adaptés indépendamment. Dans une nouvelle étude publiée en 1959, où il présente l'évolution des courbes tvi^2 en fonction de la longueur d'onde du champ d'adaptation, il conclut à l'existence de cinq mécanismes, qu'il nomme "*mécanismes pi* ". Ce résultat lui suggère que la théorie de von Kries est sans doute simpliste comme explication de l'adaptation du système visuel.

Bowmaker montrera en 1978 que les trois principaux mécanismes identifiés par Stiles sont en accord avec la courbe de sensibilité des trois types de cones humain, mais l'explication des données de Stiles reste dans le détail problématique.

²threshold-vs-intensity : la modification minimale ΔI d'une certaine variable à réaliser pour chaque valeur I de la variable afin que cette modification soit détectée.

A.3.2 La structure des similarités des couleurs

Les couleurs sont un cas particulièrement intéressant dans le domaine des similarités perceptives : l'espace des couleurs n'est pas homogène, et nous distinguons, de manière intuitive semble-t-il, les changements d'intensité lumineuse des changements de saturation ou de teinte. En 1913, Abney démontre que les notions de teinte et de saturation ne sont pas aussi triviale qu'imaginées jusqu'alors. Pour Grassmann par exemple, un changement de saturation à teinte constante est obtenu par l'ajout, à une certaine lumière, d'une proportion variable de lumière blanche. Or Abney montre qu'un tel processus change la teinte de la couleur. Cet effet vient s'ajouter à l'effet Bezold-Brücke décrit en 1874 par Johann von Bezold et Erns von Brücke : lorsque l'intensité d'une lumière monochromatique croît, l'apparence des lumières de longueur d'onde inférieure à 500 nm devient plus bleue, et celles de longueur d'onde supérieure à 500 nm deviennent plus jaune. Cet effet est généralement présenté, dans le cadre de la théorie de Hering, comme l'indication d'une interférence entre les mécanismes "rouge-vert" et "bleu-jaune". Quoiqu'il en soit, ces effets imposent une distinction entre les changements de couleur auxquels nous référons par les termes "teinte", "saturation", "luminosité", et des modifications physiques élémentaires telles que l'ajout d'une lumière blanche ou le changement d'intensité de la lumière.

En 1920, Erwin Shrödinger cherche à donner une formalisation mathématique abstraite de la structure de l'ensemble des couleurs, et argumente d'après les expériences de Maxwell que cet ensemble peut être doté d'une structure de variété tri-dimensionnelle affine. Il remarque cependant que cette structure n'implique en rien que les différences perceptives entre les couleurs doivent correspondre à une distance euclidienne, et ajoute que ces différences perceptives devraient pouvoir être mieux formulées comme une métrique riemannienne sur cette variété. Il propose que les surfaces d'isoluminance sont des ensembles de couleurs qui ont toutes les valeurs possible de saturation et de teinte, mais sont autrement minimalement distinctes localement. Il montre qu'une métrique dérivée d'une généralisation tri-dimensionnelle de la loi de Feshner, telle que suggérée par Helmholtz, donne des courbes d'isoluminance irréalistes.

William David Wright et F.H.G. Pitt, dans une série de travaux entre 1934 et 1947, s'attellent à déterminer les capacités de discrimination locale dans l'espace des couleurs. Ils indentifient entre autre la courbe t_{vi} de discrimination spectrale – une courbe décrivant, pour chaque lumière monochromatique de longueur d'onde λ , la modifi-

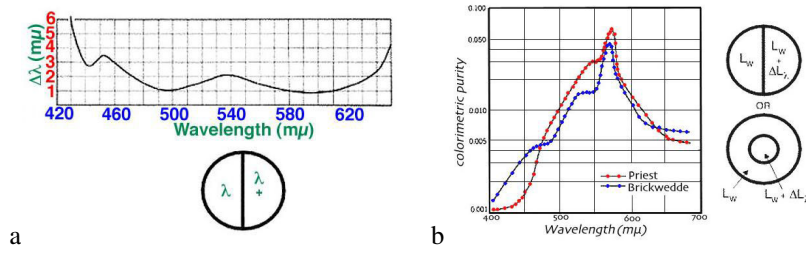


FIG. A.6 – Capacités de discriminations. a) courbes de discrimination spectrale, b) courbe de pureté

cation minimale $\Delta\lambda$ (Just Noticeable Difference) de la longueur d'onde qui produise une lumière discriminable –, et la courbe t_{vi} de pureté – une courbe décrivant, pour chaque lumière monochromatique de longueur d'onde λ , l'intensité minimale de lumière blanche à ajouter pour former une lumière discriminable (Fig. A.3.2). Selon un procédé similaire, ils identifient plus généralement les JNDs dans diverses directions sur l'ensemble d'un diagramme chromatique.

En 1942 un chercheur des laboratoires Kodak, David MacAdam, se base sur une approche légèrement différente : il utilise la dispersion des égalisations d'un observateur pour 25 points du diagramme chromatique CIE. Ces données seront d'une grande utilité pour caractériser la qualité des reproductions colorimétriques. Cependant, bien que certaines tendances générales se dégagent de l'ensemble de ces travaux, il semble que les différences inter-individuelles sont très importantes – en tout état de cause, les conditions d'adaptation du système visuel modifient la discriminabilité, ce qui n'est pas pris en compte dans ces données.

A.4 Quelques points plus récents

A.4.1 L'adoption de la théorie de l'opposition

En parallèle d'une meilleure compréhension physiologique des mécanismes rétiens (confirmation de l'existence de trois types de cones³ et de leur lien avec notre patrimoine génétique (Nathans, 1999), analyse des mécanismes de transduction moléculaire (Rushton, 1972), courbes d'absorption mesurées directement au niveau des cones (Dartnall et al., 1983; Schnapf et al., 1987), etc), la deuxième partie du XX^{eme}

³Voir tout de même Neitz et al. (1993)

siècle voit l'émergence de données et de raisonnements qui vont donner un nouveau souffle à la théorie de Hering.

En 1955, les psychophysiciens Dorothea Jameson et Leo Hurvich mesurent ce qu'ils nomment "l'annulation de teinte" (Jameson and Hurvich, 1955). Tout d'abord, ils demandent à leurs sujets d'ajuster la longueur d'onde d'une lumière monochromatique pour qu'elle leur apparaisse "sans rouge ni vert" (i.e. parfaitement jaune ou bleue), puis sans "bleu ni jaune" (i.e. parfaitement rouge ou verte), identifiant ainsi quatre lumières primaires pour le sujet. Ensuite, pour chacune des longueurs d'ondes jugées, par exemple, "avoir du rouge" par le sujet, l'intensité de la lumière primaire verte nécessaire pour que l'observateur juge que la superposition "ne contient plus de rouge" est déterminée. Jameson et Hurvich argumentent que ces données démontrent et mesurent l'équilibrage de l'activité dans deux voies rouge/vert et bleu/jaune, et qu'elles donnent une base quantitative à la théorie de l'opposition des couleurs. Ces travaux seront repris de nombreuses fois : sur la fiabilité des mesures dans les faibles longueurs d'onde (Ingling et al., 1978), sur le rapport avec le nommage des couleurs (Werner and Wooten, 1979), sur la stabilité de cette annulation vis-à-vis de l'additivité (Ayama et al., 1985; Ayama and Ikeda, 1986), sur la monotonie de la contribution de chaque cône à ces canaux théoriques (Knoblauch and Shevell, 2001), et sur la variance interindividuelle des mesures (Pugh and Larimer, 1980; Malkoc et al., 2005).

Dans les années 1960-1980, d'autre part, de nombreuses expériences psychophysiques, analysant les modifications des capacités de discrimination dans diverses conditions d'adaptation (voir Krauskopf (1999) pour une revue), sont venues renforcer l'idée de l'existence de deux principaux mécanismes distincts d'inhibition, traduisant l'existence de deux voies combinant les stimulations des cônes de manière antagoniste (voies $L - M$ et $S - (L + M)$).

A côté de ces recherches psychophysiques, les travaux de neurophysiologie sur les primates montrent qu'une partie des cellules ganglionnaires encodent l'information chromatique de manière antagoniste (Monasterio and Gouras, 1975; Livingstone and Hubel, 1984), et que les cellules du corps genouillé latéral peuvent être classées en quatre catégories selon leur réponse aux stimuli chromatiques, quatre catégories répondant préférentiellement à des stimuli approximativement rouge, vert, jaune et bleu (De Valois et al., 1958, 1967; Derrington et al., 1984; Lee et al., 1988) : les cellules du système parvocellulaire répondent à une stimulation sur l'axe rouge-vert, et les cellules

du système koniocellulaire à une stimulation sur l'axe bleu-jaune⁴. Ce type particulier de codage présenterait un avantage en termes d'optimisation du transfert d'information, en décorrélant les différents signaux transmis par la rétine (Buchsbbaum and Gottschalk, 1983). Ces résultats (malgré certaines données discordantes, e.g. Gouras and Eggers (1984)), sont alors largement tenus pour confirmer la prédiction de Hering, et pour être le pendant physiologique des précédentes données psychophysiques.

Aujourd'hui, il est généralement considéré que ces ensembles de données, qui semblent tous pointer dans la même direction, doivent en fait être distingués car ils ne coïncident pas dans le détail. Ainsi, les données psychophysiques sur l'annulation de teinte, d'une part, et sur l'adaptation des capacités de discrimination, d'autre part, suggèrent des voies neuronales combinant les signaux des cônes de manière substantiellement différente (Krauskopf, 1999; Gegenfurtner, 2003), l'adaptation des capacités de discrimination suggérant même finalement qu'il existe de nombreux mécanismes d'inhibition et pas seulement deux (Webster and Mollon, 1994). Si l'on suppose que cette inhibition a lieu à un niveau cortical et pas à un niveau thalamique ou rétinien, ce dernier résultat est cohérent avec le fait que, au niveau cortical, les cellules ne peuvent plus être classées en quatre catégories : leur sélectivité à la composition spectrale du stimulus se répartit sur un continuum (voir (Gegenfurtner, 2003) pour une revue). Enfin, la sensibilité préférentielle des quatre types de cellules observées au niveau du corps genouillé latéral distinguent en réalité des composition spectrales qui sont significativement différentes de celles singularisées par les expériences d'annulation de teinte (Valberg, 2001). Les données de Jameson et Hurvich, en fait, sont dans une large mesure la conséquence directe de l'identification de quatre "couleurs pures", indépendamment de toute théorie d'annulation (Mollon and Jordan, 1997). La question de l'existence de quatre couleurs singulières d'un point de vue phénoménologique est donc aujourd'hui une question qui reste une énigme.

A.4.2 Le débat anthropologique sur le relativisme des couleurs

Dans un domaine a priori assez éloigné de la psychophysique et de la neurophysiologie, celui de l'anthropologie, un débat général sur le rôle du langage dans la perception s'est cristallisé sur la question des couleurs. Selon une hypothèse controversée,

⁴Il faut noter que, du fait des propriétés physiologiques de ces différents systèmes, les temps caractéristiques de propagation pour ces différentes informations sont distincts.

l'hypothèse Sapir-Whorf, le langage joue un rôle critique sur la perception ([Whorf, 1964](#)) :

We dissect nature along lines laid down by our native languages. The categories and types that we isolate from the world of phenomena we do not find there because they stare every observer in the face ; on the contrary, the world is presented in a kaleidoscopic flux of impressions which has to be organized by our minds and this means largely by the linguistic systems in our minds.

En particulier, à partir d'une étude sur le langage des Inuits, Benjamin Whorf argue en 1956 que ceux-ci ont une multitude de mots pour désigner la neige et que cela influence leur perception de l'environnement⁵.

Dans de nombreux travaux du domaine, la couleur est alors tenue pour être un exemple paradigmatique de relativisme linguistique et de déterminisme du langage sur la perception. C'est dans ce contexte que deux anthropologues de Berkeley, Brent Berlin et Paul Kay, se lancent dans une étude multi-culturelle de la catégorisation linguistique des couleurs, au travers de près d'une centaine de langues. Les conclusions qu'ils publient en 1969 sont qu'il existe des universaux, que la catégorisation des couleurs dans la plupart des langues sont cohérentes entre elles. Selon eux, onze couleurs "focales" apparaissent dans la plupart des langues, et il existe une séquence typique de raffinement du vocabulaire de la couleur ([Berlin and Kay, 1969](#)). La première des conclusions obtiendra un nouveau support de la part d'une grande étude, nommée le World Color Survey (WCS), menée dans les années 90 sur les langages de cultures isolées non industrielles ([Kay and Maffi, 1999](#); [Kay and Regier, 2003](#)). Selon Berlin et Kay, l'existence d'universaux sur la catégorisation linguistique des couleurs est à mettre en relation avec la théorie de Hering, et surtout avec les résultats neurophysiologiques obtenus sur la représentation de la couleur (mais comme on l'a déjà dit, il n'y a cependant pas actuellement d'explication neurophysiologique de l'existence de couleurs "particulières").

Il faut noter, toutefois, que l'hypothèse whorfienne de la relativité linguistique garde de fervents soutiens : la reconnaissance des couleurs est influencée par le vocabulaire des couleurs, une influence qui pourrait être la véritable origine de l'aspect catégoriel

⁵L'affirmation selon laquelle les peuples arctiques ont un nombre particulièrement élevé de mots pour décrire la neige sera critiquée par le linguiste Geoffrey Pullum en 1991.

de la perception des couleurs (Davidoff et al., 1999; Roberson et al., 2000; Davidoff, 2001).

A.4.3 la constance des couleurs

L'étude des couleurs a enfin depuis les années 80 de nombreux travaux concernant la constance des couleurs – i.e. le fait que, malgré une stimulation rétinienne aussi dépendante de l'illumination de la scène que des propriétés des surfaces, notre expérience de la scène est dans une large mesure invariante lorsque l'illuminant est modifié. L'une des premières tentative d'explication de la constance des couleurs est la théorie du Retinex développée dans Land (1977). Cette théorie, basée sur l'idée d'une normalisation de la sensibilité des photorecepteurs au travers d'une adaptation continue de ces derniers pendant les mouvements oculaires, s'est ensuite effacée au profit principalement de théories basées sur des modèles linéaires des spectres lumineux et des réflectances (Brill, 1978; Maloney and Wandell, 1985; Iverson and D'Zmura, 1994; Maloney, 1999) : ces modèles s'appuient sur le fait qu'il est possible de décrire, avec une bonne approximation, le spectre des illuminants et des propriétés de réflexion des surface à partir d'un nombre restreint de fonctions de bases. Dans ces conditions, la stimulation des photorecepteurs est alors une fonction bilinéaire des illuminants et des réflectances, ce qui permet d'approcher le problème de manière géométrique.

Buchsbaum (1980) suggère que la moyenne des réflectances dans une scène est approximativement achromatique (“gray world hypothesis”), et que donc la moyenne des tristimuli est proche du tristimulus pour l'illuminant. Il propose ainsi une normalisation de la sensibilité des photorecepteurs d'une classe à partir de la moyenne des activité de cette classe.

D'Zmura and Lennie (1986) proposent une estimation de l'illuminant à partir du modèle dichromatique, en notant que ce dernier appartient à l'intersection des plans dans lesquels varient les tristimulus associés à différentes surfaces, lorsque le point de vue change. Ils proposent à partir de cette estimation une adaptation à deux niveaux : au premier niveau des photorecepteurs, et au second niveau des voies antagonistes.

Maloney and Wandell (1985) notent que dans le modèle bilinéaire, si la dimension de l'espace des réflectances est inférieur au nombre de type de photorecepteurs, les différents stimuli obtenus pour une scène sont répartis dans un sous-espace de l'espace des stimuli. Ce sous-espace détermine l'illuminant à un facteur d'échelle près, ce

qui permet d'obtenir une invariance par une transformation linéaire sur l'espace des stimuli.

D'Zmura and Iverson (1993, 1994) conduit une analyse théorique complète des modèles bilinaires, et identifie les condition d'identifiabilité des surfaces à partir de plusieurs vues, dans un modèle de dimensions arbitraires pour les réflectances, les illuminants, les photorecepteurs, et le nombre de vue.

Plus récemment, Foster and Nascimento (1994); Nascimento et al. (2002) ont suggéré qu'il n'y a pas de constance des couleurs en un sens absolu, et qu'il n'existe qu'une notion de constance des couleurs relationnelle, cette dernière étant obtenue par l'invariance des rapports des activations des photorecepteurs entre chaque couples de surfaces.

Conclusion

Dans cette thèse, nous avons proposé d'expliquer notre discours sur les sensations en termes de possibilités d'analogies entre différentes structures d'interactions sensorimotrices, analogies réalisées au travers d'adaptations sensorielles et motrices. Nous avons appliqué cette idée à notre discours sur la sensation de spatialité, et à celui sur les sensations de couleur.

Les interactions sensorimotrices dans lesquelles un organisme peut se trouver engagé sont l'ensemble des évolutions physiologiques et neurophysiologiques possibles pour cet organisme étant donnés, d'une part, la relation induite par son environnement entre ses commandes motrices et sa stimulation sensorielle (dépendance sensorimotrice), et, d'autre part, les réflexes moteurs et le contrôle rationnel de l'action induit par son fonctionnement interne (contrôle sensorimoteur). Parce que la structure de ces interactions dépend à la fois de notre système nerveux, de notre système sensorimoteur, et du monde dans lequel se trouve notre organisme, cette approche est à contraster avec la théorie, plus commune, que les sensations sont "une chose en nous", et que les données s'y rapportant doivent trouver une explication autonome dans les propriétés de notre cortex.

Nous avons tout d'abord proposé une analyse du débat scientifique autour de l'expérience perceptive, au travers de différents travaux et théories classiques en neurosciences et en psychologie de la perception. Cette analyse a permis de donner une idée des difficultés soulevées par le cadre explicatif habituellement considéré pour rendre compte des données sur les sensations, et de légitimer l'intérêt d'en explorer de nouvelles, telles par exemple que celle avancée dans cette thèse.

Dans une seconde partie, nous nous sommes concentrés sur la notion de spatialité de l'expérience perceptive. Après avoir rappelé, en particulier, les argumentations dé-

veloppées par Poincaré puis Piaget, nous avons présenté nos travaux sur le sujet, qui visent à montrer au travers de simulations que notre discours sur l'espace coïncide effectivement avec certaines propriétés de symétrie des dépendances sensorimotrices à la base de nos interactions avec l'environnement. Plus généralement, nous avons montré, en considérant le cas d'un système sensorimoteur idéalisé, que plusieurs distinctions strictes peuvent être tracées dans l'ensemble continu des dépendances induites par une variété d'environnements. D'une part, certaines dépendances présentent des invariances qui leur sont propres, et il n'est donc pas possible de les mettre en rapport avec d'autres dépendances moins symétriques. D'autre part, le système de correspondances qui existe entre différentes modifications compensables de l'environnement distingue certains s de modifications particuliers, notamment vis-à-vis de leurs propriétés de commutation. Ces propriétés conduisent à distinguer les unes des autres des modifications d'environnement telles que les mouvements rigides de la scène visuelle, et les changements d'illumination ou d'intensité sonore de l'environnement.

Dans une troisième et dernière partie, nous avons cherché à appliquer des idées similaires vis-à-vis des dépendances induites par les propriétés chromatiques des surfaces et les champs lumineux naturels. A partir du modèle dichromatique de réflexion de la lumière et d'un modèle trilineaire des illuminants naturels, nous avons montré l'existence de symétries de l'ensemble des dépendances existant entre la position relative des yeux, l'ouverture de l'iris, et les stimulations sensorielles associées des trois types de photorecepteurs rétiniens. En particulier, nous montrons que les changements de dépendances associés à un changement de luminosité ou de saturation des surfaces correspondent à des changements de dépendance compensables par des adaptations motrices du système visuel, et que les propriétés chromatiques des surfaces peuvent être distingués selon la quantité d'invariances, vis-à-vis d'adaptation sensorielles du système visuel, de la dépendance associée. La classification des surfaces suggérée par cette approche coïncide étroitement avec les données anthropologiques et psychophysiques sur la classification des couleurs.

Bibliographie

- Adrian, E. (1928). *At the Basis of Sensation*. Christopher (London).
- Ahissar, E. and Arieli, A. (2001). Figuring space by time. *Neuron*, 32 :185–201.
- Alais, D. and Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol*, 14(3) :257–262.
- Aloimonos, J., Weiss, I., and Bandyopadhyay, A. (1988). Active vision. *International Journal of Computer Vision*, 1(4) :333–356.
- Austin, J. (1962). *Sense and Sensibilia*. Oxford : Clarendon.
- Auvray, M., Hanneton, S., and O'Regan, J. K. (2007). Learning to perceive with a visuo-auditory substitution system : localisation and object recognition with 'the voice'. *Perception*, 36(3) :416–430.
- Ayama, M. and Ikeda, M. (1986). Additivity of yellow chromatic valence. *Vision Res*, 26(5) :763–769.
- Ayama, M., Kaiser, P. K., and Nakatsue, T. (1985). Additivity of red chromatic valence. *Vision Res*, 25(12) :1885–1891.
- Baars, B. J. (1993). How does a serial, integrated and very limited stream of consciousness emerge from a nervous system that is mostly unconscious, distributed, parallel and of enormous capacity ? *Ciba Found Symp*, 174 :282–90 ; discussion 291–303.
- Baars, B. J. (2005). Global workspace theory of consciousness : toward a cognitive neuroscience of human experience. *Prog Brain Res*, 150 :45–53.
- Bach-y Rita, P., Collins, C. C., Sauters, F., White, B., and Scadden, L. (1969). Vision substitution by tactile image projection. *Nature*, 221 :963–964.

- Baddeley, A. (1998). The central executive : a concept and some misconceptions. *J Int Neuropsychol Soc*, 4(5) :523–526.
- Bain, A. (1868). *The Senses and the Intellect*. London : Longman, Greens, and co.
- Ballard, D. H. (1991). Animate vision. *Artificial Intelligence*, 48(1) :57–86.
- Ballard, D. H., Hayhoe, M. M., Pook, P. K., and Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behav Brain Sci*, 20(4) :723–42 ; discussion 743–67.
- Barlow, H. B. (1972). Single units and sensation :a neuron doctrine for perceptual psychology ? *Perception*, 1(1) :371–394.
- Bartels, A. and Zeki, S. (2000). The architecture of the colour centre in the human visual brain : new results and a review. *Eur J Neurosci*, 12(1) :172–193.
- Bates, E., Elman, J., Johnson, M., Karmiloff-Smith, A., Parisi, D., and Plunkett, K. (1998). *A Companion to Cognitive Science*, chapter Innateness and emergentism, pages 590–601. Oxford : Basil Blackwell.
- Beer, R. D. (2000). Dynamical approaches to cognitive science. *Trends in Cognitive Sciences*, 4(3) :91–99.
- Berlin, B. and Kay, P. (1969). *Basic Color Terms : their Universality and Evolution*. Berkeley, University of California Press.
- Berthoz, A. (1997). *Le sens du mouvement*. Odile Jacob, Paris.
- Blakemore, C. and Mitchell, D. E. (1973). Environmental modification of the visual cortex and the neural basis of learning and memory. *Nature*, 241 :467–468.
- Block, N. (1995). On a confusion about a function of consciousness. *Behavioral and Brain Sciences*, 18(2) :227–287.
- Block, N. (1996). What is functionalism ? *The Encyclopedia of Philosophy Supplement*.
- Boi, L. (1995). *Le problème mathématique de l'espace. Une quête de l'intelligible*. Berlin : Springer-Verlag.

- Braddick, O., Atkison, J., Hood, B., Harkness, W., and Faraneh Vargha-Khadem, G. J. (1994). Possible blindsight in infants lacking one cerebral hemisphere. *Nature*, 360 :461–463.
- Brainard, D. H. and Freeman, W. T. (1997). Bayesian color constancy. *J Opt Soc Am A Opt Image Sci Vis*, 14(7) :1393–1411.
- Brainard, D. H., Longère, P., Delahunt, P. B., Freeman, W. T., Kraft, J. M., and Xiao, B. (2006). Bayesian model of human color constancy. *J Vis*, 6(11) :1267–1281.
- Brill, M. H. (1978). A device performing illuminant-invariant assessment of chromatic relations. *J Theor Biol*, 71(3) :473–478.
- Brooks, R. and Stein, L. (1994). Building brains for bodies. *Autonomous Robots*, 1(1) :pp. 7–25.
- Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310.
- Buchsbaum, G. and Gottschalk, A. (1983). Trichromacy, opponent colours coding and optimum colour information transmission in the retina. *Proceedings of the royal society of London. Series B, Biological sciences*, 220(1218) :89–113.
- Byrne, A. and Hilbert, D. R. (1997). *Readings on color, Volume 1 : The philosophy of color*. MIT Press.
- Calvin, W. H. and Stevens, C. F. (1967). Synaptic noise as a source of variability in the interval between action potentials. *Science*, 155(764) :842–844.
- Campos, J. J., Anderson, D. I., Barbu-Roth, M. A., Hubbard, E. M., Hertenstein, M. J., and Witherington, D. (2000). Travel broaden the mind. *Infancy*, 1(2) :149–219.
- Chalmers, D. J. (1995a). *Conscious Experience*, chapter Absent Qualia, Fading Qualia, Dancing Qualia. Imprint Academic.
- Chalmers, D. J. (1995b). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, 2(3) :200–219.
- Chalmers, D. J. (2000). *Neural Correlates of Consciousness :Empirical and Conceptual Questions*, chapter What is a Neural Correlate of Consciousness ? Neural Correlates of Consciousness :Empirical and Conceptual Issues. MIT Press.

- Chiao, C.-C., Osorio, D., Vorobyev, M., and Cronin, T. W. (2000). Characterization of natural illuminants in forests and the use of digital video data to reconstruct illuminant spectra. *Journal of the Optical Society of America A.*, 17(10) :1713–1721.
- Chrea, C., Valentin, D., Sulmont-Ross, C., Ly Mai, H., Hoang Nguyen, D., and Abdi, H. (2004). Culture and odor categorization : agreement between cultures depends upon the odors. *Food Quality and Preference*, 15 :669–679.
- Churchland, P. (1994). Can neurobiology teach us anything about consciousness ? *Proceedings and Addresses of the American Philosophical Association*.
- Clark, A. (1999). An embodied cognitive science ? *Trends in Cognitive Sciences*, 3(9) :345–351.
- Crair, M. C., C., G. D., and Stryker, M. P. (1998). The role of visual experience in the development of columns in cat visual cortex. *Science*, 279 :566–570.
- Crick, F. and Koch, C. (1990). Some reflections on visual awareness. *Cold Spring Harb Symp Quant Biol*, 55 :953–962.
- Crick, F. and Koch, C. (2003). A framework for consciousness. *Nature Neuroscience*, 6(2) :119–126.
- Cun, Y., Kanter, I., and Solla, S. (1991). Eigenvalues of covariance matrices : Application to neural-network learning. *Physical Review Letters*.
- Dallenbach, K. M., Supa, M., and Cotzin, M. (1944). "facial vision" the perception of obstacles by the blind. *American Journal of Psychology*, 57 :133–83.
- Dartnall, H. J., Bowmaker, J. K., and Mollon, J. D. (1983). Human visual pigments : microspectrophotometric results from the eyes of seven persons. *Proc R Soc Lond B Biol Sci*, 220(1218) :115–130.
- Davidoff, J. (2001). Language and perceptual categorisation. *Trends Cogn Sci*, 5(9) :382–387.
- Davidoff, J., Davies, I., and Roberson, D. (1999). Colour categories in a stone-age tribe. *Nature*, 398 :203–204.

- De Valois, R., Abramov, I., and Mead, W. (1967). Single cell analysis of wavelength discrimination at the lateral geniculate nucleus in the macaque. *Journal of Neurophysiology*, 30(3) :415–33.
- De Valois, R., Smith, C., Kitai, S., and Karoly, A. (1958). Response of single cells in monkey lateral geniculate nucleus to monochromatic light. *Science*, 127(3292) :238–9.
- Dehaene, S. and Naccache, L. (2001). Towards a cognitive neuroscience of consciousness : basic evidence and a workspace framework. *Cognition*, 79(1-2) :1–37.
- Deneve, S., Latham, P. E., and Pouget, A. (1999). Reading population codes : a neural implementation of ideal observers. *Nat Neurosci*, 2(8) :740–745.
- Dennett, D. C. (1992). *Consciousness explained*. Back Bay Books.
- Derrington, A. M., Krauskopf, J., and Lennie, P. (1984). Chromatic mechanisms in lateral geniculate nucleus of macaque. *J Physiol*, 357 :241–265.
- Desimone, R. and Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci*, 18 :193–222.
- D’Zmura, M. and Iverson, G. (1993). Color constancy. i. basic theory of two-stage linear recovery of spectral descriptions for lights and surfaces. *J Opt Soc Am A Opt Image Sci Vis*, 10(10) :2148–2165.
- D’Zmura, M. and Iverson, G. (1994). Color constancy. iii. general linear recovery of spectral descriptions for lights and surfaces. *J Opt Soc Am A Opt Image Sci Vis*, 11(9) :2398–2400.
- D’Zmura, M. and Lennie, P. (1986). Mechanisms of color constancy. *Journal of the Optical Society of America A*, 3(10) :1662–1672.
- Edelman, G. M. (2003). Naturalizing consciousness : a theoretical framework. *Proc Natl Acad Sci U S A*, 100(9) :5520–5524.
- Edelman, S. and Intrator, N. (1997). Learning as extraction of low-dimensional representations. In Medin, D., Goldstone, R., and Schyns, P., editors, *Mechanisms of Perceptual Learning*. Academic Press.

- Epstein, W., Hughes, B., Schneider, S., and Bach-y Rita, P. (1986). Is anything out there ? : a study of distal attribution in response to vibrotactile stimulation. *Perception*, 15 :275–284.
- Feldman, J. A. (1985). Four frames suffice : A provisional model of vision and space. *Behavioral & Brain Sciences*, 8 :265–289.
- Fellous, J. M., Houweling, A. R., Modi, R. H., Rao, R. P., Tiesinga, P. H., and Sejnowski, T. J. (2001). Frequency dependence of spike timing reliability in cortical pyramidal cells and interneurons. *J Neurophysiol*, 85(4) :1782–1787.
- Fendrich, R. and Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Percept Psychophys*, 63(4) :719–725.
- Findlay, J. M. and Gilchrist, I. D. (2003). *Active Vision : The Psychology of Looking and Seeing*. Oxford Univ. Press.
- Fiorani, M., Rosa, M. G., Gattass, R., and Rocha-Miranda, C. E. (1992). Dynamic surrounds of receptive fields in primate striate cortex : a physiological basis for perceptual completion ? *Proc Natl Acad Sci U S A*, 89(18) :8547–8551.
- Foster, D. H. and Nascimento, S. M. (1994). Relational colour constancy from invariant cone-excitation ratios. *Proceedings of the Royal Society of London B.*, 257 :115–121.
- Frith, Perry, and Lumer (1999). The neural correlates of conscious experience : an experimental framework. *Trends Cogn Sci*, 3(3) :105–114.
- Gegenfurtner, K. R. (2003). Cortical mechanisms of colour vision. *Nature reviews*, 4 :563–572.
- Geisler, W. S., Perry, J. S., Super, B. J., and Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Res*, 41(6) :711–724.
- Gerstner, W. (2000). Population dynamics of spiking neurons : fast transients, asynchronous states, and locking. *Neural Comput*, 12(1) :43–89.
- Gibson, J. J. (1979). *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates.

- Gibson, J. J. (1992). *A Century of Psychology as Science*, chapter Conclusions from a Century of Research on Sense Perception. Washington D.C : American Psychological Association.
- Gilbert, C. D., Sigman, M., and Crist, R. E. (2001). The neural basis of perceptual learning. *Neuron*, 31(5) :681–697.
- Goldstone, R. L. (1998). Perceptual learning. *Annu Rev Psychol*, 49 :585–612.
- Gomperz, T. (1908-1910). *Les penseurs de la Grèce : histoire de la philosophie antique*. Paris : F. Alcan.
- Gouras, P. and Eggers, H. (1984). Hering's opponent colour channels do not exist in the primate retinogeniculate pathway. *Ophthalmic Res*, 16(1-2) :31–35.
- Grassmann, H. (1854). On the theory of compound colors. *Philosophical Magazine*, 4(7).
- Griffin, D. and Kahneman, D. (2002). *Heuristics and Biases : The Psychology of Intuitive Judgment*. Cambridge University Press.
- Griffiths, T. D., Uppenkamp, S., Johnsrude, I., Josephs, O., and Patterson, R. D. (2001). Encoding of the temporal regularity of sound in the human brainstem. *Nat Neurosci*, 4(6) :633–637.
- Grimes, J. (1996). *Perception : Vancouver studies in cognitive science*, volume 5, chapter On the failure to detect changes in scenes across saccades, pages 89–110. Oxford : Oxford University Press.
- Hameroff, S. (2001). Consciousness, the brain, and spacetime geometry. *Ann NY Acad Sci*, 929 :74–104.
- Harris, C. M. and Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature*, 394(6695) :780–784.
- Harris, C. M. and Wolpert, D. M. (2006). The main sequence of saccades optimizes speed-accuracy trade-off. *Biol Cybern*, 95(1) :21–29.
- He, S., Cavanagh, P., and Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature*, 383(6598) :334–337.

- He, S. and MacLeod, D. I. (2001). Orientation-selective adaptation and tilt after-effect from invisible patterns. *Nature*, 411(6836) :473–476.
- Held, R. and Hein, A. (1963). Movement-produced stimulation in the development of visually guided behavior. *J Comp Physiol Psychol*, 56 :872–876.
- Henderson, J. M. and Hollingworth, A. (1999a). High-level scene perception. *Annu Rev Psychol*, 50 :243–271.
- Henderson, J. M. and Hollingworth, A. (1999b). The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10 :438–443.
- Hernandez-Andres, J., Romero, J., Nieves, J. L., and Lee, R. L. J. (2001). Color and spectral analysis of daylight in southern europe. *Journal of the Optical Society of America A*, 18(6) :1325–1335.
- Hirsch, H. and Spinelli, D. (1970). Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science*, 168 :869–871.
- Hurley, S. and Noë, A. (2003). Neural plasticity and consciousness : Reply to block. *Trends Cogn Sci*, 7(8) :342.
- Ingling, C. R., Russell, P. W., Rea, M. S., and Tsou, B. H.-P. (1978). Red - green opponent spectral sensitivity : Disparity between cancellation and direct matching methods. *Science*, 201(4362) :1221–1223.
- Irwin, D. E. (1991). Information integration across saccadic eye movements. *Cognit Psychol*, 23(3) :420–456.
- Iverson, G. and D’Zmura, M. (1994). Criteria for color constancy in trichromatic bilinear models. *Journal of the Optical Society of America A*, 11(7) :1970–1975.
- Jackson, F. (1986). What mary didn’t know. *The Journal of Philosophy*, LXXXIII(5) :291–295.
- James, W. (1981/1890). *The Principles of Psychology*. Harvard University Press.
- Jameson, D. and Hurvich, L. M. (1955). Some quantitative aspects of an opponent-colors theory. i. chromatic responses and spectral saturation. *Journal of the Optical Society of America*, 45(8) :546–552.

- Jonides, J., Irwin, D. E., and Yantis, S. (1983). Failure to integrate information from successive fixations. *Science*, 222(4620) :188.
- Judd, D. (1966). Fundamental studies of color vision from 1860 to 1960. *Proceedings of the National Academy of Sciences*, 55(6) :1313–1330.
- Kaiser, P. K. and Boynton, R. M. (1996). *Human Color Vision*. Optical Society of America.
- Kandel, E. R., Schwarz, J. H., and Jessel, T. M. (2000). *Principles of Neural Science*. McGraw-Hill.
- Kanwisher, N. and Yovel, G. (2006). The fusiform face area : a cortical region specialized for the perception of faces. *Philos Trans R Soc Lond B Biol Sci*, 361(1476) :2109–2128.
- Kaplan, F. and Hafner, V. (2005). Mapping the space of skills : An approach for comparing embodied sensorimotor organizations. In *Proceedings of the 4th IEEE International Conference on Development and Learning (ICDL-05)*, pages 129–134. IEEE.
- Kaufman, L. (1974). *Sight and Mind : An Introduction to Visual Perception*. Oxford University Press.
- Kay, P. and Maffi, L. (1999). Color appearance and the emergence and evolution of basic color lexicons. *American Anthropologist*, 101 :743–760.
- Kay, P. and Regier, T. (2003). Resolving the question of color naming universals. *Proceedings of the National Academy of Sciences USA*, 100(15) :9085–9.
- Kelemen, C. and Szirmay-Kalos, L. (2001). A microfacet based coupled specular-matte brdf model with importance sampling. In *Eurographics*.
- Kelly, S. D. and Murray, R. M. (1994). Geometric phases and robotic locomotion. Technical report, California institute of technology.
- Köhler, W. (1938). *The place of value in a world of facts*. Liveright (New York).
- Kleiser, R., Seitz, R. J., and Krekelberg, B. (2004). Neural correlates of saccadic suppression in humans. *Curr Biol*, 14(5) :386–390.

- Knoblauch, K. and Shevell, S. K. (2001). Relating cone signals to color appearance : failure of monotonicity in yellow/blue. *Vis Neurosci*, 18(6) :901–906.
- Koch, C. and Crick, F. (2004). *The Visual Neurosciences*, chapter The Neuronal Basis of Visual Consciousness, pages 1682–1694. MIT Press.
- Koch, C. and Segev, I. (2000). The role of single neurons in information processing. *Nature Neuroscience*, 3 :1171–1177.
- Koenderink, J. J. (1984). Geometrical structures determined by the functional order in nervous nets. *Biol Cybern*, 50(1) :43–50.
- Koenderink, J. J., van Doorn, A. J., Kristin, D. J., and Nayar, S. (1999). Bidirectional reflection distribution function of thoroughly pitted surfaces. *International Journal of Computer Vision*, 31(2/3) :129–144.
- Komatsu, H. (2006). The neural mechanisms of perceptual filling-in. *Nat Rev Neurosci*, 7(3) :220–231.
- Korn, H. and Faber, D. S. (1990). Transmission at a central inhibitory synapse. iv. quantal structure of synaptic noise. *J Neurophysiol*, 63(1) :198–222.
- Korn, H. and Faure, P. (2003). Is there chaos in the brain ? ii. experimental evidence and related models. *C R Biol*, 326(9) :787–840.
- Krauskopf, J. (1999). *Color Vision : From Genes to Perception*, chapter Higher order color mechanisms, pages 303–17. New York : Cambridge Univ. Press.
- Kulikowski, J. J. and Tolhurst, D. J. (1973). Psychophysical evidence for sustained and transient detectors in human vision. *J Physiol*, 232(1) :149–162.
- Lambert, J. (1760). *Photometria Sive de Mensura et Gradibus Luminus, Colorum et Umbrae*. Eberhard Klett.
- Lamme, V. A. and Roelfsema, P. R. (2000). The distinct modes of vision offered by feedforward and recurrent processing. *Trends Neurosci*, 23(11) :571–579.
- Lamme, V. A. F. (2004). Separate neural definitions of visual consciousness and visual attention ; a case for phenomenal awareness. *Neural Netw*, 17(5-6) :861–872.
- Land, E. H. (1977). The retinex theory of color vision. *Sci Am*, 237(6) :108–128.

- Lee, B. B., Martin, P. R., and Valberg, A. (1988). The physiological basis of heterochromatic flicker photometry demonstrated in the ganglion cells of the macaque retina. *J Physiol*, 404 :323–347.
- Levine, J. (1983). Materialism and qualia : the explanatory gap. *Pacific Philosophical Quarterly*, 64 :354–361.
- Liversedge and Findlay (2000). Saccadic eye movements and cognition. *Trends Cogn Sci*, 4(1) :6–14.
- Livingstone, M. S. and Hubel, D. H. (1984). Anatomy and physiology of a color system in the primate visual cortex. *J Neurosci*, 4(1) :309–356.
- Locke, J. (1690). *An Essay Concerning Human Understanding*. London.
- Logothetis, N. K. (1998). Single units and conscious vision. *Philos Trans R Soc Lond B Biol Sci*, 353(1377) :1801–1818.
- Lotze, H. (1894). *Microcosmus : An Essay Concerning Man and His Relation to the World*. T & T Clark Edinburgh.
- Macaluso, E., Frith, C., and Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. *Science*, 289 :1206–1208.
- MacKay, D. M. (1967). *Models for the perception of speech and visual form*, chapter Ways of looking at perception. MIT Press.
- Mainen, Z. F. and Sejnowski, T. J. (1995). Reliability of spike timing in neocortical neurons. *Science*, 268(5216) :1503–1506.
- Malkoc, G., Kay, P., and Webster, M. A. (2005). Variations in normal color vision. iv. binary hues and hue scaling. *J Opt Soc Am A Opt Image Sci Vis*, 22(10) :2154–2168.
- Maloney, L. T. (1999). *Color Vision :From Genes to Perception*, chapter Physics-based approaches to modeling surface color perception. Cambridge University Press.
- Maloney, L. T. (2001). Illuminant cues in surface color perception :tests of three candidate cues. *Vision Research*, 41 :2581–2600.
- Maloney, L. T. and Wandell, B. A. (1985). Color constancy :a method for recovering surface spectral reflectance. *Journal of the Optical Society of America A*.

- Mamassian, P. and Landy, M. S. (2001). Interaction of visual prior constraints. *Vision Res*, 41(20) :2653–2668.
- Marr, D. (1982). *Vision :a computational investigation into the human representation and processing of visual information*. W. H. Freeman, San Francisco.
- Marsden, J. E. and Ostrowski, J. (1998). Symmetries in motion :geometric foundations of motion control. *Nonlinear Sci. Today*.
- Martinez-Conde, S., Macknik, S. L., and Hubel, D. H. (2004). The role of fixational eye movements in visual perception. *Nature reviews Neuroscience*, 5 :229–240.
- Maye, J., Werker, J. F., and Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82(3) :B101–B111.
- McConkie, G. W. and Currie, C. B. (1996). Visual stability across saccades while viewing complex pictures. *J Exp Psychol Hum Percept Perform*, 22(3) :563–581.
- McGurk, H. and MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, 264(5588) :746–748.
- Meredith, M. A., Nemitz, J. W., and Stein, B. (1987). Determinants of multisensory integration in superior colliculus neurons. i. temporal factors. *Journal of Neuroscience*, 10 :3215–3229.
- Merleau-Ponty, M. (1945). *Phénoménologie de la Perception*. Gallimard.
- Mill, J. S. (1865). *An Examination of Sir William Hamilton's Philosophy*, CW, v. 9.
- Miller, G. (2005). What is the biological basis of consciousness? *Science*, 309(5731) :79.
- Müller, J. (1826). *Handbuch der physiologie des Menschen für Vorlesungen*, 2 vols. Coblenz : Holscher.
- Mollon, J. (2003). *The Science of Color*, chapter The origins of modern color science,, pages 1–39. Elsevier.
- Mollon, J. and Jordan, G. (1997). On the nature of unique hues. In Dickinson, C., Murray, I., and Carden, D., editors, *John Dalton's Colour Vision Legacy*, pages 381–392. London, UK : Taylor & Francis.

- Monasterio, F. M. D. and Gouras, P. (1975). Functional properties of ganglion cells of the rhesus monkey retina. *J Physiol*, 251(1) :167–195.
- Moody, T. C. (1994). Conversations with zombies. *Journal of Consciousness Studies*, 1(2) :196–200.
- Munhall, K. G., Gribble, P., Sacco, L., and Ward, M. (1996). Temporal constraints on the mcgurk effect. *Percept Psychophys*, 58(3) :351–362.
- Musen, G. and Treisman, A. (1990). Implicit and explicit memory for visual patterns. *J Exp Psychol Learn Mem Cogn*, 16(1) :127–137.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, LXXXIII(4) :435–50.
- Nascimento, S. M., Ferreira, F. P., and Foster, D. H. (2002). Statistics of spatial cone-excitation ratios in natural scenes. *Journal of the Optical Society of America A*, 19(8) :1484–1490.
- Nathans, J. (1999). The evolution and physiology of human color vision : Insights from molecular genetic studies of visual pigments. *Neuron*, 24 :299–312.
- Nayar, S. K. and Oren, M. (1995). Visual appearance of matte surfaces. *Science*, 267(5201) :1153–1156.
- Neitz, J., Neitz, M., and Jacobs, G. H. (1993). More than three different cone pigments among people with normal color vision. *Vision Res*, 33(1) :117–122.
- Newcombe, N. S. (2002). The nativist-empiricist controversy in the context of recent research on spatial and quantitative development. *Psychol Sci*, 13(5) :395–401.
- Noë, A. (2005). *Action in perception*. MIT Press.
- Noë, A. and O'Regan, J. K. (2000). Perception, attention and the grand illusion. *Psyche*, 6(15).
- Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (mmn) and its magnetic equivalent (mmnm). *Psychophysiology*, 38(1) :1–21.

- Olsson, L., Nehaniv, C., and Polani, D. (2006). From unknown sensors and actuators to actions grounded in sensorimotor perceptions. *Connection Science*, 18 :121–144.
- O'Regan, J., Myin, E., and Noë, A. (2001). Towards an analytic phenomenology : The concepts of bodiliness and grabbiness. In Carsetti, A., editor, *Proceedings of the International Colloquium : "Seeing and Thinking. Reflections on Kanizsa's Studies in Visual Cognition"*. Univ. Tor Vergata, Rome.
- O'Regan, J. K. (1992). Solving the "real" mysteries of visual perception : the world as an outside memory. *Can J Psychol*, 46(3) :461–488.
- O'Regan, J. K. and Lévy-Schoen, A. (1983). Integrating visual information from successive fixations : Does trans-saccadic fusion exist ? *Vision Research*, 23 :765–768.
- O'Regan, J. K. and Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5).
- O'Regan, J. K., Rensink, R. A., and Clark, J. J. (1999). Blindness to scene changes caused by mudsplashes. *Nature*, 34 :398.
- Oren, M. and Nayar, S. K. (1996). A theory of specular surface geometry. *International Journal of Computer Vision*, 24 :105–124.
- Paillard, J. (1999). *Cognitive contribution to the perception of spatial and temporal events*, chapter Motor determinants of a unified world perception, pages 95–111. Amsterdam : Elsevier Science BV.
- Pavani, F., Spence, C., and Driver, J. (2000). Visual capture of touch : out-of-the-body experiences with rubber gloves. *Psychol Sci*, 11(5) :353–359.
- Penfield, W. (1959). The interpretive cortex ; the stream of consciousness in the human brain can be electrically reactivated. *Science*, 129(3365) :1719–1725.
- Philipona, D., O'Regan, J. K., and Nadal, J.-P. (2003). Is there something out there ? Inferring space from sensorimotor dependencies. *Neural Computation*, 15(9).
- Piaget, J. (1937). *La construction du réel chez l'enfant*. Neuchâtel ; Paris.
- Piaget, J. and Inhelder, B. (1972). *La représentation de l'espace chez l'enfant*. Presses Universitaires de France.

- Pierce, D. and Kuipers, B. (1997). Map learning with uninterpreted sensors and effectors. *Artificial Intelligence*, 92 :169–229.
- Poincaré, H. (1895). L'espace et la géométrie. *Revue de Métaphysique et de Morale*, 3 :630–646.
- Poincaré, H. (1898). On the foundations of geometry. *The Monist*, 9 :1–43.
- Poincaré, H. (1902). *La Science et l'Hypothèse*. Flammarion.
- Posner, M. I. and Petersen, S. E. (1990). The attention system of the human brain. *Annu Rev Neurosci*, 13 :25–42.
- Posner, M. I. and Rothbart, M. K. (1998). Attention, self-regulation and consciousness. *Philos Trans R Soc Lond B Biol Sci*, 353(1377) :1915–1927.
- Price, H. H. (1950). *Perception*. London : Methuen.
- Pugh, E. N. and Larimer, J. (1980). Test of the identity of the site of blue/yellow hue cancellation and the site of chromatic antagonism in the pi 1 pathway. *Vision Res*, 20(9) :779–788.
- Putnam, H. (1987). *The Many Faces of Realism*. La Salle, Open Court.
- Radford, J. and Burdick, J. (1998). Local motion planning for nonholonomic control systems evolving on principal bundles. In *Mathematical Theory of Networks and Systems*.
- Ramachandran, V. S. and Gregory, R. L. (1991). Perceptual filling in of artificially induced scotomas in human vision. *Nature*, 350(6320) :699–702.
- Rayner, K. and Pollatsek, A. (1992). Eye movements and scene perception. *Can J Psychol*, 46(3) :342–376.
- Reed, E. (1990). *James's Principles, in Reflections on the Principles of Psychology : William James after a Century*, chapter Space Perception and the Psychologist's Fallacy. New Jersey : Laurence Elbaum Associates.
- Rees, G., Kreiman, G., and Koch, C. (2002). Neural correlates of consciousness in humans. *Nature reviews Neuroscience*, 3 :261–270.

- Regan, D., Giaschi, D., Sharpe, J. A., and Hong, X. H. (1992). Visual processing of motion-defined form : selective failure in patients with parietotemporal lesions. *J Neurosci*, 12(6) :2198–2210.
- Renier, L., Laloyaux, C., Collignon, O., Tranduy, D., Vanlierde, A., Bruyer, R., and Volder, A. G. D. (2005). The ponzo illusion with auditory substitution of vision in sighted and early-blind subjects. *Perception*, 34(7) :857–867.
- Rensink, R., O'Regan, J., and Clark, J. (1997). To see or not to see : The need for attention to perceive changes in scenes. *Psychological Science*, 8(5) :368–373.
- Rensink, R. A. (2002). Change detection. *Annu Rev Psychol*, 53 :245–277.
- Riddoch, M. and Humphreys, G. (1987). A case of integrative visual agnosia. *Brain*, 110 :1431–1462.
- Roberson, D., Davies, I., and Davidoff, J. (2000). Color categories are not universal : replications and new evidence from a stone-age culture. *J Exp Psychol Gen*, 129(3) :369–398.
- Rolls, E. T., Tovée, M. J., and Panzeri, S. (1999). The neurophysiology of backward visual masking : information analysis. *J Cogn Neurosci*, 11(3) :300–311.
- Romero, J., Garcia-Beltran, A., and Hernandez-Andres, J. (1997). Linear bases for representation of natural and artificial illuminants. *Journal of the Optical Society of America A*, 14(5) :1007–1014.
- Rosenthal, D. M. (2002). How many kinds of consciousness? *Conscious Cogn*, 11(4) :653–665.
- Ross, J., Morrone, M. C., Goldberg, M. E., and Burr, D. C. (2001). Changes in visual perception at the time of saccades. *Trends Neurosci*, 24(2) :113–121.
- Roweis, S. T. and Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290 :2323–2326.
- Rushton, W. A. (1972). Pigments and signals in colour vision. *J Physiol*, 220(3) :1P–1P.
- Russell, B. (1997/1912). *The Problems of Philosophy*. New York : Oxford University Press.

- Ryle, G. (1990/1949). *The concept of mind*. Penguin Books.
- Schnapf, J. L., Kraft, T. W., and Baylor, D. A. (1987). Spectral sensitivity of human cone photoreceptors. *Nature*, 325(6103) :439–441.
- Schneider, G. E. (1973). Early lesions of superior colliculus : factors affecting the formation of abnormal retinal projections. *Brain Behav Evol*, 8(1) :73–109.
- Shallice, T. (1988). *Consciousness in contemporary science*, chapter Information-processing models of consciousness : possibilities and problems. Oxford : Clarendon.
- Sharma, J., Angelucci, A., and Sur, M. (2000). Induction of visual orientation modules in auditory cortex. *Nature*, 404(6780) :841–847.
- Shimojo, S. and Shams, L. (2001). Sensory modalities are not separate modalities : plasticity and interactions. *Current Opinion in Neurobiology*, 11 :505–509.
- Simons, D. (2000). Current approaches to change blindness. *Visual Cognition*, 7 :1–15.
- Simons, D. and Ambinder, M. (2005). Change blindness : Theory and consequences. *Current Directions in Psychological Science*, 14(1) :44–48.
- Slater, A., Mattock, A., and Brown, E. (1990). Size constancy at birth : newborn infants' responses to retinal and real size. *J Exp Child Psychol*, 49(2) :314–322.
- Stockman, A. and T., S. L. (2000). The spectral sensitivities of the middle- and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision Research*, 40(13) :1711–37.
- Sur, M., Angelucci, A., and Sharma, J. (1999). Rewiring cortex : the role of patterned activity in development and plasticity of neocortical circuits. *J Neurobiol*, 41(1) :33–43.
- Tenenbaum, J. B. (1998). Mapping a manifold of perceptual observations. *Neural Information Processing Systems*, 10.
- Tenenbaum, J. B., de Silva, V., and Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290 :2319–2323.

- Thiele, A., Henning, P., Kubischik, M., and Hoffmann, K.-P. (2002). Neural mechanisms of saccadic suppression. *Science*, 295(5564) :2460–2462.
- Valberg, A. (2001). Unique hues : an old problem for a new generation. *Vision Research*, 41 :1645–1657.
- van Vreeswijk, C. and Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science*, 274(5293) :1724–1726.
- von Helmholtz, H. (1867). *Optique physiologique, tomes I et II*.
- von Helmholtz, H. (1876). The origin and meaning of geometrical axioms. *Mind*, 1(3) :301–321.
- von Melchner, L., Pallas, S. L., and Sur, M. (2000). Visual behaviour mediated by retinal projections directed to the auditory pathway. *Nature*, 404(6780) :871–876.
- Wallace, M. T. and Stein, B. (1994). Cross-modal synthesis in the mid-brain depends on input from association cortex. *Journal of Neurophysiology*, 71 :429–432.
- Watson, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20 :158–177.
- Webster, M. A. and Mollon, J. D. (1994). The influence of contrast adaptation on color appearance. *Vision Res*, 34(15) :1993–2020.
- Weerd, P. D. (2006). Perceptual filling-in : More than the eye can see. *Prog Brain Res*, 154 :227–245.
- Weerd, P. D., Gattass, R., Desimone, R., and Ungerleider, L. G. (1995). Responses of cells in monkey visual cortex during perceptual filling-in of an artificial scotoma. *Nature*, 377(6551) :731–734.
- Wegner, D. M. (2003). The mind's best trick :how we experience conscious will. *Trends in Cognitive Sciences*, 7(2) :65–69.
- Weiskrantz, L., Barbur, J. L., and Sahraie, A. (1995). Parameters affecting conscious versus unconscious visual discrimination with damage to the visual cortex (v1). *Proc Natl Acad Sci U S A*, 92(13) :6122–6126.

- Weiss, Y., Simoncelli, E. P., and Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nat Neurosci*, 5(6) :598–604.
- Werner, J. S. and Wooten, B. R. (1979). Opponent chromatic mechanisms : relation to photopigments and hue naming. *J Opt Soc Am*, 69(3) :422–434.
- Whorf, B. L. (1964). *Language, thought and reality*. Cambridge MA : MIT Press.
- Wiggs, C. L. and Martin, A. (1998). Properties and mechanisms of perceptual priming. *Curr Opin Neurobiol*, 8(2) :227–233.
- Wittgenstein, L. (2001/1953). *Philosophical Investigations*. Blackwell Publishing.
- Wundt, W. (1886). *Eléments de psychologie physiologique - Volume 1*. Paris.
- Yang, J. N. and Maloney, L. T. (2001). Illuminant cues in surface color perception : tests of three candidate cues. *Vision Res*, 41(20) :2581–2600.
- Zeki, S. (1993). The mystery of louis verrey (1854-1916). *Gesnerus*, 50 (Pt 1-2) :96–112.
- Zeki, S. and Bartels, A. (1998). The autonomy of the visual systems and the modularity of conscious vision. *Philosophical Transactions : Biological Sciences*.
- Zeki, S. and Marina, L. (1998). Three cortical stages of colour processing in the human brain. *Brain*, 121(9) :1669–1685.

Bibliographie personnelle

Articles de journal avec comité de lecture

- [1] M. Auvray, D. Philipona, J. K. O'Regan, and C. Spence. The perception of space and form recognition in a simulated environment : The case of minimalist sensory-substitution devices. *Perception*, 36(12) :1736 – 1751, 2007.
- [2] D. Philipona, J. K. O'Regan, and J.-P. Nadal. Is there something out there ? Inferring space from sensorimotor dependencies. *Neural Computation*, 15(9), 2003.
- [3] D. Philipona, J. K. O'Regan, J.-P. Nadal, and O. J.-M. Coenen. Perception of the structure of the physical world using unknown sensors and effectors. *Advances in Neural Information Processing Systems*, 15, 2004.
- [4] D. L. Philipona and J. K. O'Regan. Color naming, unique hues, and hue cancellation predicted from singularities in reflection properties. *Visual Neurosciences*, 23(3-4) :331–339, 2006.

Chapitres et proceedings

- [5] D. Philipona and O. J. M. D. Coenen. Model of granular layer encoding in the cerebellum. In *Neurocomputing*, volume 58-60, pages 575–580, 2004. Computational Neuroscience : Trends in Research 2004.
- [6] D. Philipona and J. K. O'Regan. *Agir dans l'espace*, chapter La perception de l'espace, identification d'une faculté sensorimotrice ? Catherine Thinus-Blanc, 2004.
- [7] D. Philipona and J. K. O'Regan. Perception multimodale de l'espace. In *La philosophie de la nature aujourd'hui ? Les apports des mathématiques, de la physique et de la biologie*. MSH, 2004.
- [8] D. Philipona and J. K. O'Regan. Perception sensorimotrice de l'espace. In *AS Suppleance Arobase*. Université de Compiègne, 2004.

Accepté avec modifications, pas encore publié

- [9] F. Glanois, D. Philipona, and J. K. O'Regan. Sensorimotor reduction of dimensionality. *Neural Computation (accepted with modifications, not yet published)*, -.